

Conducting cross-national and cross-cultural surveys: papers from the 2005 meeting of the international workshop on Comparative Survey Design and Implementation (CSDI)

Harkness, Janet (Ed.)

Veröffentlichungsversion / Published Version

Konferenzband / conference proceedings

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Harkness, J. (Ed.). (2006). *Conducting cross-national and cross-cultural surveys: papers from the 2005 meeting of the international workshop on Comparative Survey Design and Implementation (CSDI)* (ZUMA-Nachrichten Spezial, 12). Mannheim: GESIS-ZUMA. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49740-1>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

GESIS

ZUMA

NACHRICHTEN

Spezial Band 12

Conducting Cross-National and Cross-Cultural Surveys

**Papers from the 2005 Meeting of the International
Workshop on Comparative Survey Design
and Implementation (CSDI)**

Janet A. Harkness (ed.)

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung.

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz* (WGL) (www.wgl.de).

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Jan van Deth

Direktor: Prof. Dr. Peter Ph. Mohler

Hausanschrift

B 2, 1
68 159 Mannheim

Postanschrift

Postfach 12 21 55
68 072 Mannheim

Telefon 0621/1246 - 184

Fax 0621/1246 - 100

E-Mail Harkness@zuma-mannheim.de

Internet www.gesis.org/zuma ZUMA

www.gesis.org/ GESIS

ISBN 3-924220-31-X

Druck: Druck & Kopie Hanel, Mannheim

© **ZUMA, Mannheim, 2006**

Preface

Janet A. Harkness.....5

What's So Special About Cross-National Surveys?

Peter Lynn, Lilli Japek & Lars Lyberg.....7

Fieldwork Details in the European Social Survey 2002/2003

Achim Koch & Michael Blohm.....21

Discussion of the Income Measure in the European Social Survey:

A Proposal of Revised Survey Questions About the "Total Net Household Income"

Uwe Warner & Jürgen H.P. Hoffmeyer-Zlotnik.....53

An Examination of Pretesting Methods for Multicultural, Multilingual Surveys:

The Use of Cognitive Interviews to Test Spanish Instruments

Patricia L. Goerman.....67

Cross-Cultural Communication and the Telephone Survey Interview

Brian Kleiner & Yuling Pan.....81

The Use of Interpreters in the Conduct of Household Surveys:

Development of U.S. Census Bureau Interpretation Guidelines

Yuling Pan.....91

Category and Comparison Across What Kind of Frontier?

John MacInnes.....101

About the Contributors.....115

PREFACE

This volume contains seven papers presented at the third meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI). The papers illustrate the broad range of methodological questions pursued by researchers active in cross-cultural and cross-national survey research.

The opening and closing papers in the volume deal with core questions in cross-national survey research. The first (*LYNN, JAEPEC & LYBERG*), written by members of a CSDI work group helping formulate a methodological research agenda for comparative research, discusses the special character and needs of cross-national research. The closing paper (*MACINNES*) focuses on challenges involved in any act of comparison and the consequences of decisions implicitly or explicitly taken in comparing given variables with one another. The second and third papers focus on cross-national research in the European context. *KOCH & BLOHM* discuss insights gained from the information on data collection available from the European Social Survey (ESS), while *WARNER & HOFFMEYER-ZLOTNIK* use ESS data on socio-demographic variables to argue their case for how to measure income in the cross-national European context. The remaining three papers in the volume deal with language issues in survey research. Problems related to language and equivalence of measurement across languages have long been favourite topics of discussion in disciplines using surveys and survey-like instruments. As researchers have turned to address the needs of minority populations within countries, language issues in nationally based survey research have also become a focus of attention. The papers included here present ongoing research in the U.S. context on translation and questionnaire pre-testing (*GOERMAN*); cross-cultural communication issues in telephone survey research (*KLEINER & PAN*); and preparations at the U.S. Census Bureau to develop guidelines on the use of interpreters in survey data collections (*PAN*).

The International Workshop on Comparative Survey Design and Implementation was initiated in 2003 as a logical international extension of cross-cultural symposia held at ZUMA through the nineteen nineties and into the present century. The Workshop provides a platform and forum for researchers working on methodological issues related to comparative survey research. CSDI activities go beyond the annual springtime meetings, however. For example, CSDI work groups have made remarkable progress recently on a range of quality monitoring and documentation issues, guidelines on translation, and research on socio-economic variables. A large international conference is also in the planning for 2008. Interested readers can find more information on CSDI activities at the CSDI website (<http://www.csdi-workshop.org>).

WHAT'S SO SPECIAL ABOUT CROSS-NATIONAL SURVEYS?

PETER LYNN¹, LILLI JAPEC & LARS LYBERG

1 Introduction

1.1 Objectives and Structure of this Paper

For the past fifteen years we have seen an increase in the number of studies comparing countries, regions, and cultures on different dimensions. This is a trend that we predict will continue. As a consequence, we are likely to see even more cross-national surveys in the future. In this paper we use the term cross-national surveys to represent all types of surveys where efforts are made to achieve comparability across countries. Efforts to achieve comparability vary on a wide spectrum from opportunistic adjustment of data after they have been collected to deliberate design of each step in the survey process to achieve functional equivalence.

Cross-national surveys differ from national surveys in a number of important ways. In this paper we will highlight some of the differences and discuss the implications for survey design and implementation. A key difference is the goal of cross-national surveys, since the focus is on comparability. This requires careful consideration of what is meant by comparability between countries (Van de Vijver, 2003; Johnson, 1998). We thus begin with a discussion of the objectives of cross-national surveys (section 2). We then organise our description of the important aspects of cross-national surveys into unique aspects of survey design (section 3), aspects related to survey planning and organisation (section 4), variation in national constraints (section 5), magnification of cultural issues (section 6), unique aspects of standardisation (section 7), control of design and implementation (section 8) and analysis and reporting (section 9). We conclude by discussing how the survey community might progress from this initial identification of issues that are particular to cross-national surveys towards a situation where we are better equipped to design and implement surveys in ways that are more likely to result in comparable data and analyses.

1 We are grateful to workshop participants for helpful comments and discussions.

1.2 Examples of cross-national surveys

Cross-national surveys cover a wide range of topics from economic indicators to literacy. Examples include: the European and World Values Surveys (EVS, WVS), the Afro barometer, the Eurobarometer, the Latinobarometer, the Harmonized European Time Use Survey (HETUS)/ Multinational Time Use Study (MTUS), the Harmonized Indices of Consumer Prices, the International Social Survey Programme (ISSP), the European Community Household Panel Survey (ECHP), EU Statistics on Income and Living Conditions (EU-SILC) and the European Social Survey (ESS). Links to websites for the cross-national surveys discussed in this section are provided in the annex.

Cross-national surveys also vary greatly in the methods used to achieve comparability across countries and cultures. Some surveys specify only a very limited set of survey design constraints. Perhaps one extreme is where only the question wordings are specified (and only in one language, with no guidance on translation), along perhaps with a minimum sample size. Other aspects of survey design are left entirely to the discretion of the participating institution in each country. The WVS takes an approach similar to this. There are clear advantages to such minimal specification. The workload on the central organisation is minimised and the likelihood of countries being able and willing to participate may be maximised.

Also close to this extreme we have European official statistics that cover many countries, cultures and languages with minimal specification. The need for statistics within the EU was recognized early on and Eurostat was set up in 1951, as part of the European Commission, to provide the EU with statistics from the member states. National Statistical Institutes (NSI) in each country carry out surveys and then Eurostat harmonizes the statistics to make them more comparable across countries. Johnson (2003) defines harmonization as

“a method for equating conceptually similar but operationally different variables that are collected as part of separate surveys for purposes of cross-cultural or cross national research. Also referred to as ‘ex post harmonization’.”

The harmonization efforts include activities such as working group meetings with survey representatives from member countries discussing issues concerning their specific survey. Discussion topics typically include definitions, standards and guidelines. Sometimes the harmonization efforts result in regulations that state, in more detail, what data should be delivered to Eurostat e.g., there is a number of regulations for the Harmonized Indices of Consumer Prices. The survey design, methods and data collection are usually left to each NSI to decide upon. The result of this is that different methods are used in a survey e.g., the EU Labour Force Survey (LFS) is carried out as a face-to-face survey in most countries but in some countries it is a telephone survey.

In a continuing national survey we would be very hesitant to change the survey methodology, e.g., the data collection mode, without a methodological study to assess the effects of the change. The fact that the choice of methods can have an impact on survey results is well documented in the survey literature. However in international surveys, especially where harmonization is the main method for making statistics comparable, we often assume ignorable methodological effects despite the fact that methodological studies showing ignorable effects are very rare. EU-SILC is a recent development where the attempts to achieve comparability go beyond the traditional harmonization efforts to include guidelines on how data should be collected.

Another area where a number of cross-national surveys are carried out is in the field of literacy and learning. Examples of surveys are Trends in Maths and Science Study (TIMSS), Programme for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS) and International Adult Literacy Survey (IALS). These surveys are all deliberately designed to achieve comparability between countries. The IALS measures adults' literacy skills given their formal schooling and reading practices in daily life (OECD 1997). An extensive effort is put into developing tests in each country so that the tests become comparable. However less effort is put into other parts of the design. For example, sampling, estimation, and the implementation of the survey are left to each country.

The European Social Survey (ESS), which collects data on, e.g., trust in politics and religion, satisfaction with government, schools and health care, is at the other end of the scale of cross-national surveys in terms of specification and control. Large efforts are made to achieve comparability. This includes not only detailed specification of design and implementation, but also processes of control, guidance and liaison. Some examples of measures taken in the ESS to achieve comparability are the development of guidelines, creation of a central sampling team that helps out with sampling and estimation in all participating countries, organisation of a translation taskforce that provides guidance on translation procedures, insistence that face to face interviews are carried out in all countries, that interviewers provide information about the interview, that call record data are collected in a standardised way, and that information is collected about issues discussed in the media in each country during the actual data collection period (Stoop, 2003).

2 Objectives of Cross-National Surveys

Goals in cross-national surveys are different from goals in national surveys. Even though data from cross-national surveys are used on national levels, one main use is to compare countries and regions (groups of countries) on different dimensions. The types of comparisons typically made are (1) comparisons of estimates of parameters for different countries or regions or (2) rankings of countries on different dimensions. The parameters in question could be averages, totals, measures of distribution, or measures of association between variables such as model coefficients, for example. Another use of data from cross-national surveys is simply (3) to aggregate estimates from a number of countries in order to provide an estimate relating to a supra-national region such as, for example, the European Union. Each of these three types of use of cross-national survey data obviously requires data that are comparable.

We can not expect meaningful comparisons if we do not take into account the specifics of the countries. To illustrate this we build on an example from Braun & Mohler (2003) about divorce rates in Europe. If the researcher ranks countries (2) on this dimension, Ireland will end up having a very low divorce rate compared to other countries in Europe. The conclusion that should be made from this is not necessarily that the Irish people are happier in their marriages than other Europeans. Instead the low divorce rate can be explained by the fact that the religion in Ireland does not allow divorce but does allow separation. It might be less obvious that we also need to think of comparability when comparing aggregates for a region (3). To illustrate this we continue our example of divorce rates. If we compare divorce rates in Europe with divorce rates in North America, we have to aggregate data for countries within each of these two regions. The same issue of comparability is still present, i.e., we need to make a decision on how to handle separation and divorce rates in these regions, and whether we should aggregate only divorce rates or include separation rates in our measure.

In other words, for cross-national surveys country is the key variable that defines analysis domains or acts as a covariate. All surveys have some important analysis domains and it is always important to ensure comparability between domains. But the difference with cross-national surveys is that one variable is always far more important than the others, and it is always the same one: country. This is the reason that we should pay particular attention to comparability *between* countries, though of course we should additionally pay attention to comparability between sample elements (and therefore groups thereof) *within* countries.

Van de Vijver (2003) discusses different aspects of comparability and equivalence. He points out that cultural specifics in comparative studies constitute a potential problem if care is not taken to distinguish these from more universal aspects. This means that we

need to ascertain which aspects are universal and which are culture specific. Having done that, we should aim to develop measurement tools that provide unbiased estimates of the *relevant* concepts. This requires valid and reliable instruments (*realism*) and unbiased *representation* with respect to the population distributions of all survey measures. The latter can only be achieved by avoiding coverage, sampling and non-response bias. Additionally, it is necessary to be able to estimate the variance of survey estimates. Meaningful comparisons between countries cannot be made if the precision of estimates is unknown. We do not develop further here the discussion of how equivalence might be defined, but recognition of the importance of the concepts of relevance, representation and realism should inform identification of relevant issues in the sections that follow.

3 Unique Aspects of Survey Design

Cross-national surveys can be considered to have an extra layer of survey design, in addition to the aspects that must be considered for any survey carried out in a single country. The first crucial component is to decide which countries are included. For some surveys, this is relatively uncontroversial, such as with European official statistics which must cover all EU member states. But even here there can be discussions about the inclusion of accession and candidate countries and other important European countries such as Switzerland, Norway and Iceland. Some surveys impose no geographical constraints on participation (e.g. WVS), or no constraints within a particular region (e.g. ESS, Afrobarometer), with participation often dependent on national funding, capacity and enthusiasm. Others identify a limited set of countries for inclusion, perhaps based on a combination of substantive interests and practical constraints. However, the set of countries included is important as it can have great influence on findings, for example regarding between-country variation, or where regions of the world or other supra-national aggregates are to be compared based on only a sample of the countries in each region or aggregate (e.g. see Brown et al., forthcoming).

The second component of design that is unique to cross-national surveys is the choice of how to distribute the sample over countries. Often, reflecting the recognition of countries as key analysis domains and between-country differences as key estimates, the choice is to aim for equal sample sizes in each country. In fact, the precision of such estimates is likely to be maximised by attempting to achieve approximately equal *effective* sample size per nation, where this takes into account the differential effects of design features such as sample clustering, stratification and variation in selection probabilities on precision. (Although this is of course by no means an easy thing to control – Lynn et al., 2004.) Some surveys, such as ESS and EU-SILC, specify national sample size requirements in terms of effective sample size.

Alternatively, some analysts may prefer to view nation as an important covariate, the aim of comparative analysis being to identify the extent and nature of variation that is explained by nation. From this perspective, it can be argued that it is the total variation in the group of nations under study that is the dependent variable and that the sample should be distributed in proportion to each nation's contribution to the total variation. The latter perspective is one of identifying variation over nations in general, rather than differences between specific nations. These two perspectives provide very different design conclusions if the nations under study vary greatly in size (as they typically do).

Whichever perspective is used, a further possible refinement is that important analysis domains may be identified within some or all countries and that sample size requirements may be specified with respect to those domains. If those domains are geographical, there may be more of them in some countries than others, leading to variation in national sample sizes where no variation would have been implied by the requirements for national estimates alone. EU-SILC is an example of a survey where this kind of refinement has been applied. However, the question of sample size per country is one that is often answered on pragmatic or resource grounds, rather than in reaction to the statistical requirements of the survey.

The third component of design that has special characteristics in the case of cross-national surveys is the identification of meaningful relevant concepts and items to study. The concepts and items must be relevant cross-nationally and are consequently often not the same ones that would be chosen for a national survey on the same topic. Both the criteria for choosing concepts and items and the processes by which the choices are made are typically very different than in the case of national surveys.

4 Survey Planning and Organisation

It is rare that cross-national surveys are carried out by a single survey agency. Typically, different survey organisations will carry out the field work in each nation, each with different ideas and working practices. It is therefore necessary for there to be an entity in charge of the entire planning and implementation phase, separate from the survey agencies. The responsibilities of such an entity can vary from simply providing instructions to individual countries to actually assisting and supervising the individual survey organizations involved (see section 8). Experience tells us that the involvement from such a central entity must be extensive if equivalence is to be achieved. Otherwise national considerations will sometimes take precedence over survey design requirements, and there will be an increased risk of simple misunderstanding of the instructions.

Ambitious cross-national studies are enormous undertakings. The design, quality control, analysis, and dissemination are on completely different levels as compared with a national survey. This has implications for almost every stage of the survey process. Considerable resources are required and skillful resource allocation is needed. Communication and liaison between multiple agencies (who will often have different working languages) also becomes a major factor in the success of a cross-national survey.

5 Variation in National Constraints

Methodological and financial resources often differ extensively between countries, as can the experience and capabilities of survey agencies. This can impose considerable constraints on the nature and extent of standardization that can be insisted upon or achieved. At the margins this can lead to trade-off decisions between relaxing one or more requirements for a particular country (with the disadvantage of possibly lesser comparability, as well as the risk of setting a precedent for permitting relaxations of the rules) or forcing the exclusion of the country from the survey (with the disadvantage of not being able to include the country in any analyses).

Legal aspects might affect what it is possible to achieve in a given country. As laws are inherently national (or even in some cases sub-national), this creates differences between countries in the types of design and implementation features that are possible. Relevant laws can include those which relate to data linkage, access to potential sampling frames or auxiliary data, informed consent, rights of privacy and freedom from harassment or unsolicited approaches.

There can be big differences between nations in sampling constraints, such as the nature of available sampling frames and legislation governing their use. This inevitably leads to a choice between permitting and dealing with controlled variation in sample design or imposing a lowest-common-denominator sample design suitable for use in all participating countries, but probably not the best possible approach in many countries.

6 Magnification of Cultural Issues

There is always a variation among individuals in a population regarding their perceptions of questions and concepts. This variation is due to differences in education, experience, cultural norms, language, sensitivity to topic and other personal traits. What happens is that the outcome of the response process is affected by this variation and questions must be worded and administered in such a way that this variation is minimized. In cross-national surveys, in addition to this within-country variation there is also between-country

variation reflecting more profound, systematic, differences between countries. Examples of such differences are that the meaning of concepts, topic sensitivity, perception of response scales, interviewer-respondent interaction, survey climate, computer access, and telephone penetration rate might vary considerably between countries. Basically one has to develop a questionnaire and a mix of data collection modes that is so robust that differences of this kind do not have a devastating effect on the comparability. In essence this means that questions should have the same meaning and be interpreted in a similar fashion across countries (and modes, when necessary).

In addition, there is typically no longer one dominant culture and language amongst the target population. Instead, several very different cultures may be of approximately equal importance to the survey (in terms of their representation in the sample(s) and impacts on survey estimates). And differences in culture (and language) make it impossible to use identical instruments in each country. In consequence, developing and testing cross-national questionnaires becomes a very complex task.

To complicate matters further all survey materials including questions and questionnaires must be translated. This is not only expensive and time-consuming, but can also have a big impact on the measurement characteristics of survey items. There is evidence that traditional translation procedures, such as back-translation, have serious drawbacks and must be replaced by more efficient and effective procedures.

7 Unique Aspects of Standardisation

Typically, national surveys are thought of as having a single “design”. By this, we mean that one sample is selected (to a specified design), one set of survey instruments are used (with a particular design), the field work is organised as one unified operation, and so on. Standardisation is applied (or at least strived for) at several levels. Each sample element receives a standard treatment to encourage participation, is administered a standard questionnaire in a standard way, by interviewers who have received the same training and instructions, and so on. With cross-national surveys, there are several reasons why this is not the case.

These include the variation in national constraints discussed above in section 5, and differences in language and culture that make it impossible to use identical instruments in each country. There is therefore pressure for differences in survey design (Lynn, 2003). It is rarely possible to ensure that every aspect of survey design is replicated in identical fashion in every nation. Neither is it necessarily desirable to do this. Rather, the challenge is to identify which aspects of design need to be identical, which should be allowed (encouraged) to vary – and within what parameters – and which may be less important, in the

sense that relevant characteristics of the survey data may be insensitive to variations in design. At present there is insufficient research evidence to provide much guidance on this question. We simply do not know the relative effects of variation in different aspects of design on the equivalence of cross-national data.

If the objective of a survey is to achieve equivalence in each nation (Jowell, 1998), for some aspects of design this may best be achieved by standardisation of design. For example, equivalence of measurement may best be achieved by administering an item in the same mode, using the same wording and same visual stimuli in every nation (though it should not be assumed that this will always be the case). On the other hand, there are some dimensions for which equivalence may best be achieved by conscious variation in design. For example, to achieve representation of an equivalent population by a sample of equivalent precision, may require the use of different sampling frames and different selection methods in different nations (Lynn et al., 2004; Adams & Wu, 2002 chapter 4). However, even in these two areas of measurement and sampling there is a paucity of knowledge about the effects of different design options on equivalence for different types of survey measures. For other aspects of survey design and implementation, such as field work practices, almost nothing is known.

The extent to which it is possible to standardise the design may depend as much on the infrastructure and processes for coordination and control (see section 8 below) as on statistical considerations. For example, with respect to measurement instruments it may be recognised that the instruments should ideally be pretested in each participating nation and in each language in which they will be administered. And that the findings of the pretests should be discussed collaboratively and collective decisions made about revisions to the instruments. But resource and logistical constraints may prevent this approach from being implemented, whereas other design aspects may be more standardised because it is easier to do so, even though they may be less important.

8 Control of Design and Implementation

On any survey, the design and the data collection and processing procedures – once decided and agreed – require control to ensure that they are carried out as intended. Survey organisations have processes and mechanisms by which to implement this control as routine, often involving regular reporting and liaison with the single survey funder. But with a cross-national survey there are typically several survey organisations and a separate co-ordinating organisations, and possibly other funding organisations too. This requires special control procedures. Without them, there is a strong risk that the design and implementation plan will not be followed adequately.

Different models have been used for the co-ordination and control of survey implementation. At one extreme, it is possible to set up a large central co-ordination team who are able to liaise closely with, and monitor the activities of, each national team throughout the implementation process. At the other extreme, the central co-ordinator may simply issue some written instructions on implementation and leave each nation to follow the instructions. The first of these two models is obviously resource intensive and requires substantial funding for the central activities.

In practice, the co-ordination model adopted may be influenced by the nature of survey funding. If each nation is funding its own participation, there is likely to be pressure for the national funds to be spent on country-specific activities, rather than contributing to a central pot. There may or may not be a separate funding source for centralised activities. However, the choice of co-ordination model is likely to have important implications for the degree of supervision and quality control that is possible. For example, the analysis of regular process data during the field work period is an important tool for monitoring adherence to certain aspects of the survey specification. If the central team is unable to perform this analysis, gross failures to adhere to the specification (which could be either wilful or accidental) may not be identified until long after the survey is completed, if ever (e.g. Park & Jowell, 1997).

As with any survey, the monitoring and control process should feed back into the survey process in a cycle of continuous quality improvement wherever this appropriate. The distinctive aspect of cross-national surveys is that this is likely to involve multiple agencies and be a relatively slow process.

9 Analysis and Reporting

The “distance” between the user and the producer is obviously greater in cross-national surveys compared with national surveys. This puts additional pressure on developing adequate documentation as the analyst will often be reliant upon the documentation to understand relevant aspects of the design and implementation.

Local knowledge (familiarity with national circumstances and contexts) is crucial as input both in the design phase and in the analysis phase. It is particularly difficult to draw informed inference from cross-national comparisons without an even and informed knowledge of the relevant context in each nation. This can be very difficult to obtain and typically requires collaboration of researchers in several nations.

There is a risk that some participating countries might challenge the results from cross-national surveys. One reason is that in ranking studies the outcome is inevitably that some

countries will be the lowest ranked. In other comparative studies specific countries might be described in a less than favorable light. National pride is obviously at stake and the typical scenario is that representatives from countries that do not do well in the studies question methods or survey organizations involved. Therefore analysis and documentation are especially crucial in cross-national surveys in order to demonstrate the scientific credentials of the survey.

10 Concluding Discussion

Much methodological development regarding cross-cultural survey methods (Harkness et al., 2003) concerns question development and translation, how to achieve equivalence in different dimensions, how to perform secondary analysis, and the effect of cultural bias. As mentioned these issues become very complex in cross-national survey settings. There is considerable activity in these methodological and conceptual areas. However, other aspects are receiving less attention and here we end our discussion by adding a few thoughts on methodological issues that might benefit from some research in a cross-national survey context.

As mentioned (section 2), there is no definition of the ideal design properties of cross-national surveys. Without an approximate definition of ideal design properties it is very difficult to know how best to allocate resources. It seems as if input harmonization, as we know it, can take us only so far. And post hoc harmonization is even more limited. We need to develop a cross-national survey methodology that emphasizes management issues and other issues that become especially problematic in cross-national surveys. The survey community is currently learning how to manage large cross-national surveys such as the European Social Survey. It would be useful to assemble the management experiences from recent efforts to form a consensus on how to run these kinds of surveys. Without decent management we will not get decent measurements.

Based on the ideal design properties, we then need to establish design principles that can guide us in the planning and trade-off situation. Advancing input harmonization from merely design specifications to a management system that includes planning, cooperation, training, quality control, assistance, analysis, documentation and continuous improvement seems to be a logical next step. The literature on so-called house effects shows that sometimes survey organizations sharing a common data collection effort produce significantly different results (Smith, 1982). The encouraging fact is that the same literature also point to data collections where no house effects have been discovered (Cohen & Potter, 1990), which means that it is possible for individual survey organizations to sometimes collect comparable data. Cross-national surveys resemble a situation with a potential for exten-

sive house effects, i.e., survey organizations in different countries are supposed to perform identical tasks in such a way that result differences between countries cannot be attributed to the survey organization's standard practices and performance. Thus, these house effects should be minimized.

Translation of survey materials is emerging as a discipline in its own right. Current standard practices such as back-translation seem to be problematic. A more efficient translation procedure is to use team translation. Guidelines need to be developed, implemented and disseminated.

It is widely agreed that good product quality can only be achieved by using underlying processes that are free from unnecessary variation, i.e. stable and predictable. One way of keeping track of the processes is through the choice, measurement and analysis of key process variables (Couper & Lyberg, 2005). Cross-national survey processes need specific sets of process data so that the complex production processes can be controlled. Also it might be worthwhile to explore the concept of responsive designs (Groves & Heeringa, 2004). Here process data are used to inform cost and quality tradeoff decisions in real time. The ability to monitor both process data and regular survey data creates an opportunity to alter the design during the survey data collection in order to improve survey cost efficiency and achieve more accurate estimates.

Any survey, national or cross-national, faces problems regarding general design issues, how to allocate resources effectively, how to perform quality control activities, and how to analyze and disseminate the results to the users. There is no comprehensive theory of survey planning and management and most survey work is based on a mixture of theories, design principles resulting from experiments and experiences, and constraints in terms of costs, time and other user demands. Of course, in a cross-national survey these general problems remain and new ones appear as well. We have attempted to identify in this paper the likely sources of such new problems. However, we know relatively little about their effect on comparability. The identification of such properties is a research priority. When more is known about the relationship between the various unique characteristics of cross-national surveys identified in this paper and between-country comparability, the research community will be in much stronger position to understand how best to design and implement cross-national surveys.

References

- Adams, R., & Wu, M. (Eds). (2002). *PISA 2000 Technical Report*. Paris: Organisation for Economic Co-operation and Development.
- Braun, M., & Mohler, P. (2003). Background Variables. In J. A. Harkness & F. J.R. Van de Vijver (Eds.), *Cross-Cultural Survey Methods*. New York: Wiley.
- Brown, G., Micklewright, J., Schnepf, S., & Waldman, R. (forthcoming) *Cross-national surveys of learning achievement: How Robust are the Findings?*
- Cohen, S., & Potter, D. E.B. (1990). Data Collection Organization Effects in the National Medical Expenditure Survey. *Journal of Official Statistics*, 6(3), 275-293.
- Couper, M., & Lyberg, L. (2005). *The Use of Paradata in Survey Research*. Invited paper presented at the 56th ISI Session, Sydney, April.
- Groves, R. M., & Heeringa, S. G. (2004). *Responsive Design for Household Surveys: Tools for Actively Controlling Survey Nonresponse and Costs*. University of Michigan, unpublished paper.
- Harkness, J., Van de Vijver, F. J.R., & Mohler, P. Ph. (2003). *Cross-Cultural Survey Methods*. Wiley.
- Johnson, T. (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. In J. A. Harkness (Ed.), *ZUMA-Nachrichten Spezial No. 3. Cross-Cultural Survey Equivalence*. Mannheim: ZUMA.
- Johnson, T. (2003). Glossary. In J. A. Harkness & F. J. R. Van de Vijver (Eds.), *Cross-Cultural Survey Methods*. New York: Wiley.
- Jowell, R. (1998). How Comparative is Comparative Research? *American Behavioural Scientist*, 42(2), 168-177.
- Lynn, P. (2003). Developing Quality Standards for Cross-National Survey Research: five approaches. *International Journal of Social Research Methodology*, 6(4), 323-337.
- Lynn, P., Gabler, S., Häder, S., & Laaksonen, S. (2004). *Methods for Achieving Equivalence of Samples in Cross-National Surveys*. Working Paper 2004-09 of the Institute for Social and Economic Research. Colchester: University of Essex.
- OECD (1997). *Literacy Skills for the Knowledge Society*. Further Results from the Adult Literacy Survey.
- Park, A. & Jowell, R. (1997). *Consistencies and Differences in a Cross-National Survey*. London: SCPR.
- Smith, T. (1982). House Effects and the Reproducibility of Survey Measurements: A Comparison of the 1980 GSS and the 1980 American National Election Study. *Public Opinion Quarterly*, 46, 54-68.

- Stoop, I. (2003). *Collecting Event Data*. Presentation at the Comparative Survey Design and Implementation Workshop in Brussels. (<http://www.gesis.org/Mitarbeiter/zuma/harkness/csdi>)
- Van de Vijver, F. J.R. (2003). Bias and Substantive Analysis. In J. A. Harkness & F. J.R. Van de Vijver (Eds.), *Cross-Cultural Survey Methods* (pp. 207-233). New York: Wiley.

Links to websites

- Afrobarometer. <http://www.afrobarometer.org/>
- Eurobarometer. http://europa.eu.int/comm/public_opinion/index_en.htm
- European Community Household Panel Survey. <http://epunet.essex.ac.uk>
- European Official Statistics. Eurostat. <http://europa.eu.int/comm/eurostat/>
- European Social Survey. <http://www.europeansocialsurvey.org/>
- European Value Survey. <http://www.europeanvalues.nl/>
- International Adult Literacy Survey. <http://www.statcan.ca/english/Dli/Data/Ftp/ials.htm>
- International Social Survey Programme. <http://www.issp.org>
- Latinobarometer. <http://www.latinobarometro.org/>
- Multinational Time Use Survey. <http://www.iser.essex.ac.uk/mtus/>
- Programme for International Student Assessment. <http://www.pisa.oecd.org/>
- Trends in Maths and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). <http://timss.bc.edu/>
- World Values Survey. <http://wvs.isr.umich.edu/>

FIELDWORK DETAILS IN THE EUROPEAN SOCIAL SURVEY 2002/2003

ACHIM KOCH & MICHAEL BLOHM

1 Introduction

The objective of the survey interview is to collect information about the population under study in a uniform and reliable way (Weinberg, 1983: 329ff). For every survey data collection is a critical part of the survey process. In contrast to other parts of this process, like questionnaire development e.g., data collection often is not in the hands of the researcher. In order to implement a face-to-face survey, the researcher usually has to cooperate with a survey organisation, and fieldwork consists of decentralised operations of lots of interviewers. That makes it often difficult to gain insights in what is actually happening in the field. Therefore, fieldwork sometimes has been called to be a “black box”. This assertion holds for a survey in a single country. It holds even more, if we turn to cross-national surveys, where several countries and survey organisations are involved. Here, quite often only aggregate information about issues like the number of interviews realized, the length of the fieldwork period, etc. are available.

The present paper tries to go a step beyond this. It provides more detailed information on several aspects of fieldwork in the European Social Survey (ESS). The ESS is a new multi-country biennial survey, which follows very high methodological standards and aims to improve survey methods and documentation of the whole survey process. In the years 2002 and 2003, the first Round of ESS was fielded as a face-to-face survey in 22 European countries.

The paper starts with a brief introduction of the ESS in section 2. In section 3 we provide some basic data on fieldwork in ESS Round 1 (number of achieved interviews, length of fieldwork period, number of interviewers). In section 4 we proceed with information on various more detailed aspects of the work of the interviewers. The main focus is on the actual interviewing task of the interviewers, in particular the temporal aspects of their

work.¹ The issues analysed include the average number of interviews realized per day; the day and time when interviews are made; the length of the start-up phase until interviewers do complete their first interview; and the length of the time period interviewers are actually engaged in realizing interviews. With this information available, we can detect characteristic differences and similarities in the work of the interviewers across countries. Do these reported fieldwork details matter? In section 5 we do some exemplary correlational analyses. First, we examine whether there is a relationship between the fieldwork details described and the length of the fieldwork period in the ESS countries. Second, we investigate, whether there is a correlation between the timing of the interviews and data quality, concrete: the degree of underrepresentation of people in paid work in the realized sample. In the conclusions in section 6 the need to replicate these analyses with other data sets is emphasized.

2 The European Social Survey

The European Social Survey (ESS) is a new multi-country biennial survey with two main aims (for more details, see <http://www.europeansocialsurvey.org/>). First, it seeks to measure, monitor and interpret changing public attitudes within Europe and the way in which they interact with Europe's changing institutions. Second, it seeks to advance and consolidate improved methods of cross-national quantitative measurement within Europe and beyond.

The project is directed by a Central Coordinating Team, led by the Centre for Comparative Social Surveys at City University, London. In each participating country, a National Coordinator is responsible for the conduct of the national survey to a common standard. The work of the central team is mainly funded by the European commission, whereas the fieldwork and other national costs in each country were to be borne by national funding agencies.

The first round of ESS was fielded in the years 2002/2003 in 22 nations. The questionnaire contained several broad topics, like public trust in government and politicians, political interest and participation, socio-political orientations, moral and social values, national, ethnic and religious allegiances, well-being, health and security, and information on social structure. The average interview length was around 70 minutes.

1 We do not cover the tasks of contacting and motivating target persons to participate, and we do not deal with issues like achieved response rates, non-contact and refusal rates in ESS Round 1 (see e.g. Philippens & Billiet, 2004).

A standard specification for all participating countries established the methods and procedures that had to be followed. The specification covered a wide range of issues from sampling, questionnaire design, event and context measuring, translation, fieldwork standards, response rates and archiving. For instance, it was laid down that all countries had to use strict random sampling methods, and that no substitution of ‘non-contacts’ or ‘refusals’ was permitted. The population should consist of all persons aged 15 and over resident within private households. A target response rate of at least 70% should be aimed at. The fieldwork should last for at least one month within a four-month period between 1 September and end December 2002. All interviews had to be carried out face-to-face. The interviewers had to be trained in face-to-face briefing sessions, and the interviewer workload was strictly limited. These specifications aimed to ensure that the most rigorous methodologies were used in the countries and to secure consistency and comparability of the resulting data. Adherence to them was a requirement for a countries’ data to be included in the integrated ESS dataset.

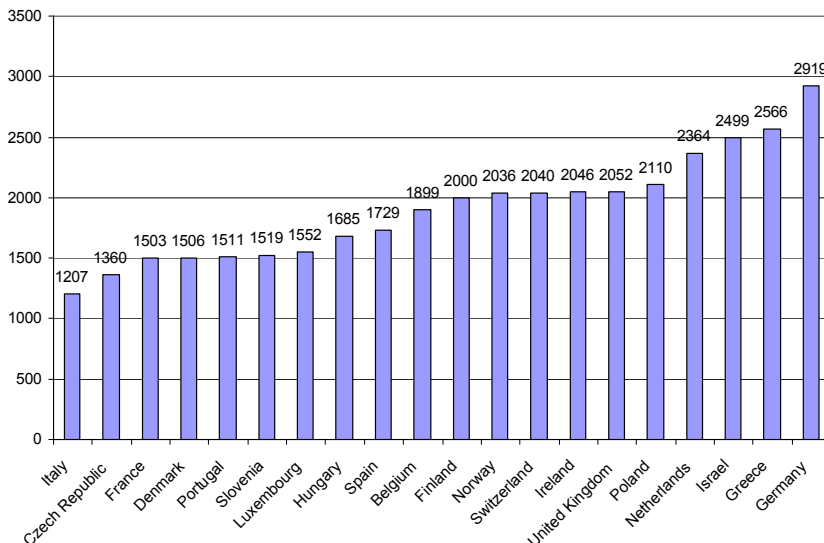
The following sections provide some insights on how the countries actually implemented data collection in ESS Round 1. Our analyses cover 20 of the 22 countries participating in the first Round of ESS. Austria and Sweden are not included, since both countries did not provide an interviewer identification number in their data sets.

3 Basic Information on Fieldwork in ESS Round 1

3.1 Number of achieved interviews

The “Specifications for participating countries” in ESS Round 1 set down a minimum sample size of 2,000 interviews in each country. Countries with populations of under two million were required to achieve at least 1,000 interviews. In Round 1, this applied only to Slovenia and Luxembourg.

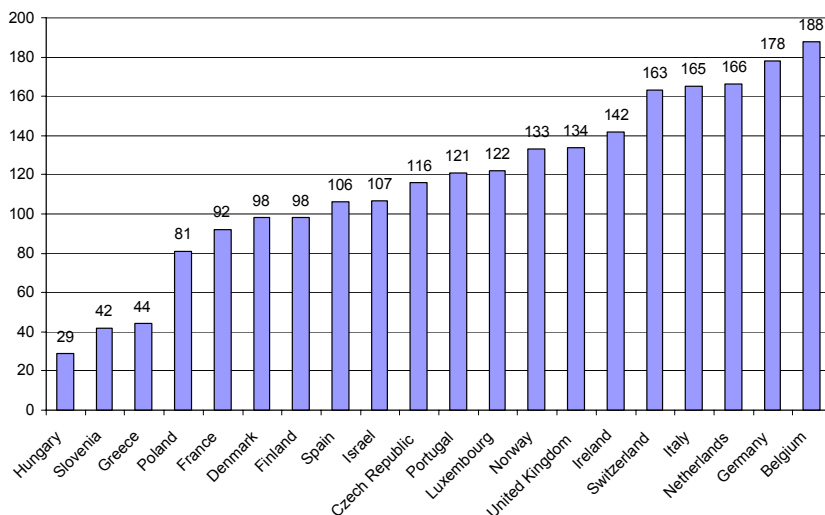
Twelve of the 20 countries met the sample size requirement (see Figure 1). Most interviews were made in Germany (2,919 interviews). Eight countries did not achieve the required number of interviews. In some countries the sample size was lower because of budgetary constraints, in other countries low response rates brought the number of interviews below 2,000. On average, 1,905 interviews were realized per country.

Figure 1 **Number of achieved interviews in ESS Round 1**

3.2 Length of fieldwork period

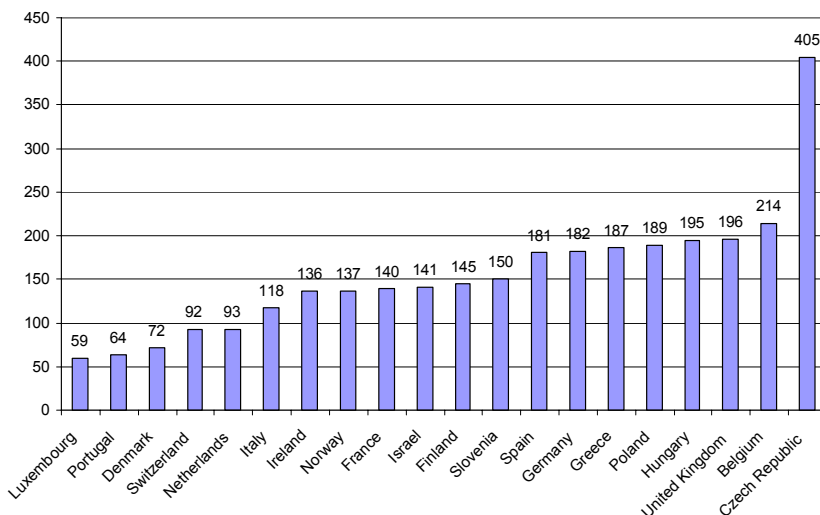
According to the “Specifications for participating countries” fieldwork in ESS Round 1 should last for at least one month within a four-months period from September to December 2002. The actual length of the fieldwork period in the countries of ESS Round 1 varied between one and somewhat more than six months (see Figure 2).² Averaging across all 20 countries the fieldwork lasted 116 days. On the one hand there were three countries with a rather short fieldwork period: In Hungary, Slovenia, and Greece fieldwork was finalized within one and one and a half month, respectively. On the other hand eight countries exceeded the required maximum fieldwork length of four months. The longest fieldwork period pertained to Germany and Belgium, where around six months were necessary.

2 With respect to the start date of fieldwork even greater variations did occur: Seven countries managed to get under way in September 2002, and another nine countries began in the course of the remaining months of 2002. Four countries started fieldwork in 2003, the last country commenced in September 2003. Difficulties in receiving the necessary national funding were the primary reason for these delays.

Figure 2 Length of fieldwork period in ESS Round 1 (in days)

3.3 Number of interviewers and number of interviews per interviewer

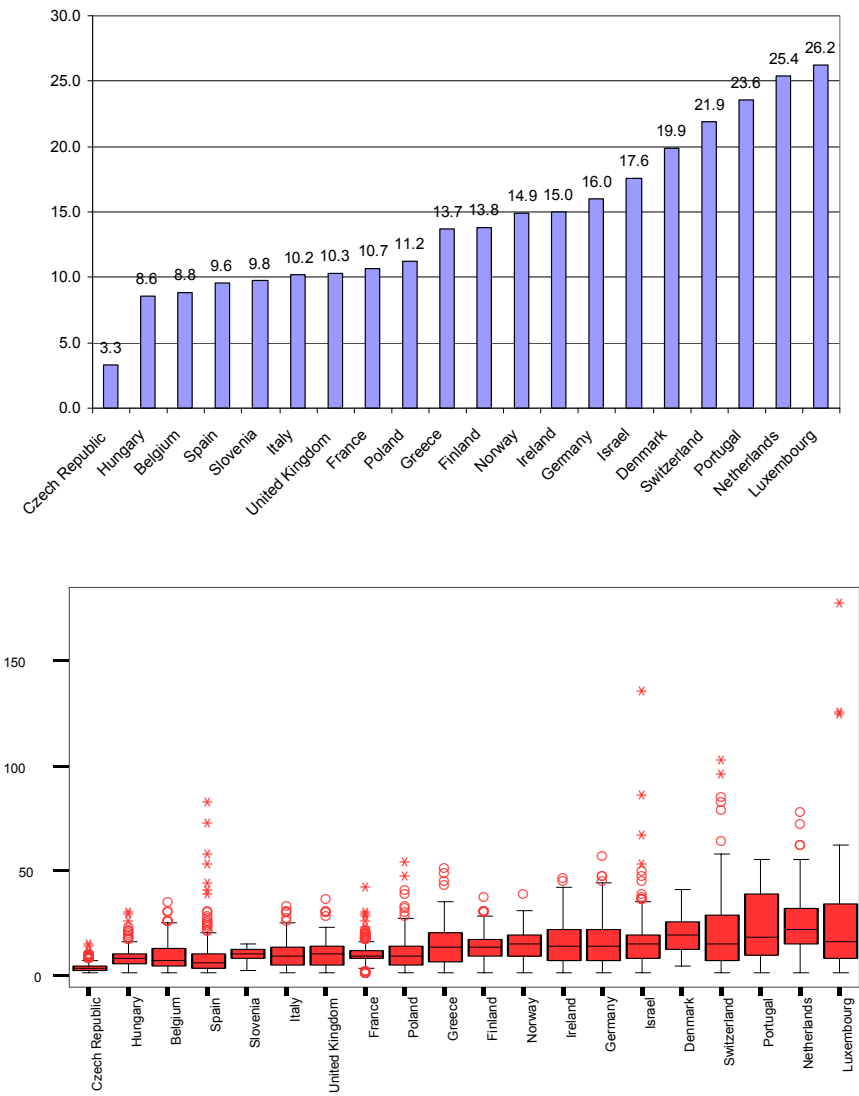
How many interviewers were involved in fieldwork for ESS Round 1? In the “Specifications for participating countries” there was no explicit requirement with regard to the number of interviewers to be deployed. Figure 3 shows, that the number of interviewers used for fieldwork in ESS 1 varied between a low of about 60 to a high of 405. On average, 155 interviewers were involved in each country. Most countries deployed between around 140 to 200 interviewers. Only one country used much more than 200 interviewers: the Czech Republic with 405 interviewers. Fewest interviewers were deployed in Luxembourg (59 interviewers) and Portugal (64 interviewers).

Figure 3 **Number of interviewers in ESS Round 1**

In determining the number of interviewers to be deployed for a study various issues play a role: The total number of interviews to be realized, the average time needed to realize an interview, the planned length of the fieldwork period, the geographical distribution of the target persons, the availability and location of the interviewers (where do they live in relation to the respondents?), the complexity of the study and the resulting needs for (personal) interviewer training,³ just to name a few of them.

3 Probably it did not happen by chance that in ESS Round 1 Czech Republic was the only country in which the interviewers were not personally briefed, since the personal briefing of all 405 interviewers would have been very expensive.

Figure 4 **Number of interviews per interviewer in ESS Round 1**
(means and boxplots of distribution)



From a methodological point of view one should keep the number of interviewers high and the average workload of the interviewers as low as possible in order to reduce the possibility of interviewer effects. "...When interviewers affect the answers they get, the impact on estimates of standard errors is directly related to the average number of interviews per interviewer" (Groves et al., 2004: 296). Using more interviewers on a particular study, and having them each take fewer interviews, is a way to reduce this effect. On the other hand one has to recognize that the interviewer workload should not be too small in order to be able to organize fieldwork efficiently. The interviewers need to do a couple of interviews to become familiar with the survey instrument; training efforts are only cost efficient, when interviewers complete at least a certain number of interviews.

In the "Specifications for participating countries" in ESS Round 1 a rule was formulated that no single interviewer should work on more than 48 individuals, households, or addresses (gross). That meant, that an interviewer should be allowed to realize 48 interviews at most (net). As can be seen in Figure 4 the average number of completed interviews per interviewer was well below this level. Across all countries, 15 interviews were completed on average, but there was some variation between countries. The interviewer workload was highest in the Netherlands and Luxembourg with an average number of 25 and 26 completed interviews. The lowest workload is observed in Czech Republic with three completed interviews on average. The boxplots in Figure 4 reveal, that in several countries some interviewers exceeded the allowed limit of 48 completed interviews. In the maximum, an interviewer in Luxembourg took 178 interviews – these add up to 11% of all interviews realized in that country.

After having provided information on these basic fieldwork issues, we will now turn to some more detailed and less commonly used indicators of fieldwork patterns (see e.g. Koch, 2002; Schnell, 1997: 229ff).

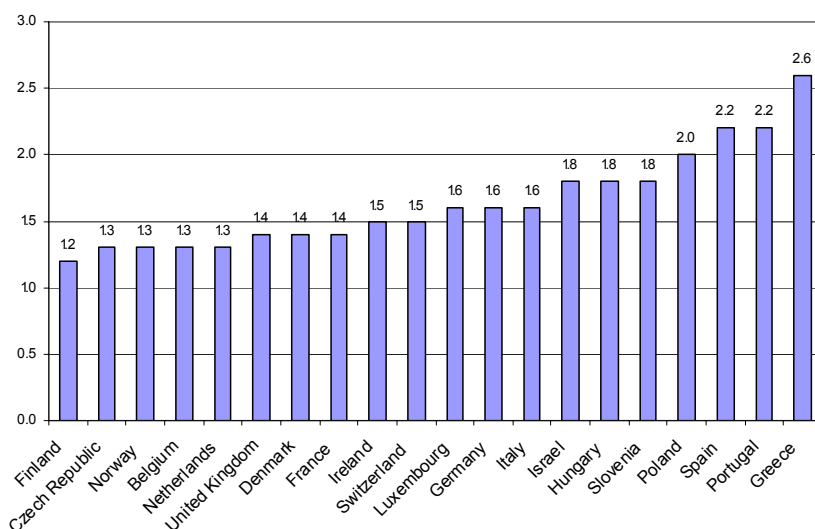
4 Fieldwork Details in ESS Round 1

4.1 Number of interviews per day per interviewer

When interviewers are working for a study like the ESS, how many interviews do they usually realize per day? Several factors will have an impact on this. First and foremost, this will depend on the time budget the interviewers devote to the interviewing task. It is

probably safe to say, that interviewing is most-often a part-time job. This limits the number of interviews to be completed on a single day. Besides the interviewer's time budget, the length of the interview,⁴ the travelling time needed to get to the home of the target persons, and the accessibility of the sample respondents (that means their at-home-patterns and their willingness to agree with the survey request) can have an influence on the number of interviews realized per day.

Figure 5 **Average number of interviews per day per interviewer in ESS Round 1**

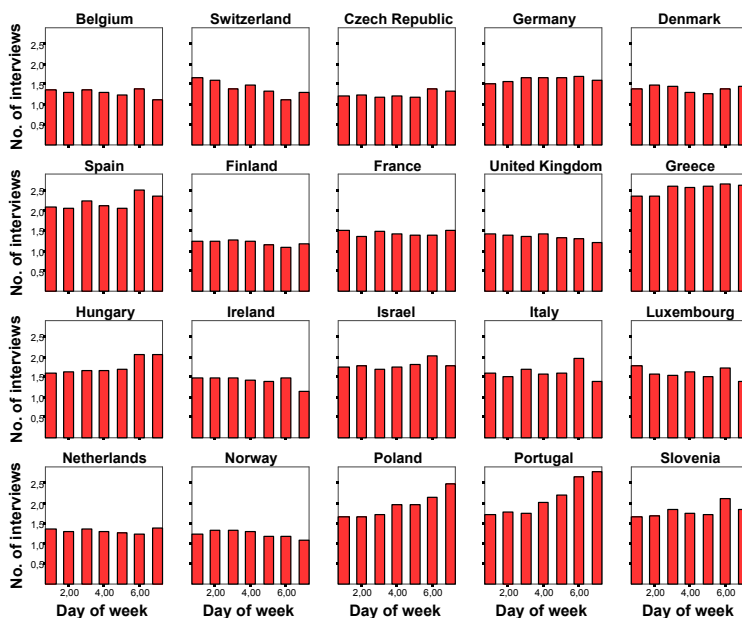


4 In ESS Round 1 the average interview length added up to 63 minutes, with some variation across countries. The shortest length was observed in Italy and Spain with 53 minutes. The interview lasted longest in Poland with 70 minutes on average. Since the additional supplementary questionnaire took around 10 minutes to be filled in, the "pure" interview time needed for a single interview amounted to about 75 minutes on average. Obviously, the ESS was not a short interview. If, for instance, an interviewer was working only in the evening, it depended on the travelling time needed to get from one target person to another whether one or two interviews were feasible at the same day.

In ESS Round 1, the interviewers completed on average 1.6 interviews per day. In 65% of the 23947 days at which interviewers were successfully working for ESS only one interview was realized. In another 22% of days two interviews were made. And in 13% of all interview days, three or more interviews were completed. There is some variation in this across countries (see Figure 5). The lowest figures are obtained for Finland (with an average of 1.2 interviews), Belgium, Czech Republic, the Netherlands, and Norway (1.3 interviews). In four countries, two or more interviews have been completed on average. These are Poland (2.0 interviews), Portugal and Spain with 2.2 interviews, and Greece (2.6 interviews). The figure for Greece is probably somewhat exaggerated, because – according to the Greek National Coordinator – in a few cases two interviewers were working under the same interviewer identification number.

Obviously, interviewing for ESS Round 1 was not a full-time job for the interviewers. This holds, even if we take into account, that a 75 minutes face-to-face interview takes more interviewer time, particularly because of the travelling time needed (Weinberg, 1983: 336f). When we look at the results separately for the different days of the week we find some variation between weekdays and weekends in a few countries. In particular in Hungary, Poland, and Portugal, the average number of completed interviews is higher on weekends than on weekdays (by 0.4 to 0.8 interviews, see Figure 6). Switzerland is showing the reverse pattern. Here, the average number of completed interviews per day is 0.3 lower on weekends compared to weekdays.

Figure 6 Average number of interviews per day per interviewer, separately for different days of the week, ESS Round 1*



* Very small sample sizes ($n < 30$) were observed for 'Sunday interviews' in Finland ($n=27$), France ($n=4$), Netherlands ($n=10$), Norway ($n=23$), Switzerland ($n=17$).

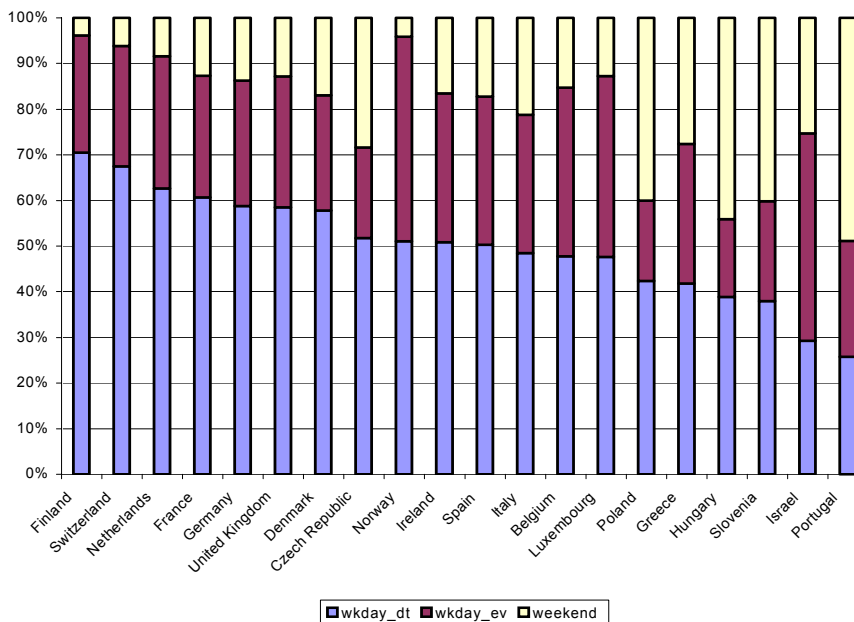
4.2 Time and day of interviews

When do the interviewers take their interviews? Are there specific times of the day and/or specific days of the week which are particularly important for interviewing? In the "Specifications for participating countries", recommendations were given regarding the number, mode and timing of call attempts. At least four personal call attempts should be made at every sample unit. At least one of these call attempts should be made in the evening and one at the weekend. The call attempts should be spread over at least 14 days.

But, in contrast, there was no explicit requirement regarding the timing of the interviews. Common sense would expect that in a survey of the general population like the ESS a considerable part of the work should be made in the evening or at weekends in order to interview those who are at work during the day.

When we define all interviews starting at 5 pm or later as evening interviews we obtain the results shown in Figure 7. On average across all countries 50% of the ESS interviews were realized on weekday morning or afternoon. 29% of the interviews were completed on weekday evening and 21% on weekends.

Figure 7 Day and time of interviews in ESS Round 1 (in %)



In Finland and Switzerland more than two thirds of all interviews were realized during weekday daytime.⁵ In contrast, in Israel and Portugal less than 30% of all interviews were completed at weekday morning or afternoon. In Israel, Norway, and Luxembourg a lot of interviews were made in the evening (40% to 46%). In Portugal, Hungary, Slovenia, and Poland between 40% and 49% of all interviews were taken on the weekend. So, only in every second country participating in ESS Round 1 the majority of interviews was realized in the evenings or on weekends, i.e. at times which seem to be most productive for household interviewing.

We know too little to draw firm conclusions about the reasons for these differences across countries. One might speculate whether the employment conditions of the interviewers play a role. In the Scandinavian countries, for instance, the respective National Statistical Agency was in charge of doing the fieldwork for ESS. In these cases the interviewers were regular employees of the survey organization. Maybe this is the reason for the low proportion of interviews taken on the weekend.⁶ The same factor may account for the rather low weekend figures in the Netherlands and Luxembourg (8% and 13%), too. In contrast, high figures for weekend interviews may result from the fact that in some countries working as an interviewer is a second job in addition to being fulltime employed at another job. So, a lot of the interviewing work has to be done on weekends.

Alternatively, the particular distribution of interviewing times in a country could result from a specific call schedule the survey organisation had implemented for ESS. According to the information in the National Technical Summaries, the interviewers in Poland and Slovenia, for instance, were required to make two contact attempts at weekends (instead of only one as prescribed by the Specifications). Both countries exhibit a high proportion of weekend interviews. In both countries, and in Israel, too, in addition two visits (instead of one) were required to be made in the evening. But with respect to the proportion of weekday evening interviews, only Israel shows a result above average.

4.3 Start-up phase until interviewers complete their first interview

After the official start of fieldwork it usually takes some days until the interviewers actually realize their first interview. Once they have received their fieldwork assignments and the study materials, like questionnaires, showcards, brochures, and contact forms, the

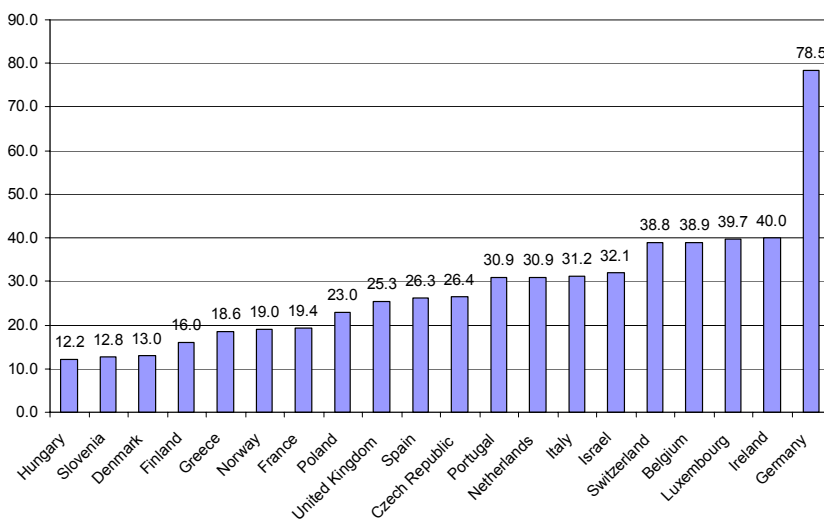
5 An even higher figure is obtained for Sweden, which is not included in our analyses: Here, 77 % of all interviews were made on weekday morning or afternoon.

6 Sweden showed an even lower figure: only 1.3% of all interviews were made on the weekend.

interviewers can begin to work. First of all they have to locate the sampled addresses and to plan the order of their visits to minimize travelling. Subsequently, they can try to contact the target persons, in order to realize their first interview or to get at least an appointment for an interview. In ESS Round 1, the interviewers completed their first interview on average four weeks after fieldwork had started.

Some countries are well below this average (see Figure 8). In Hungary, Slovenia and Denmark the interviewers realized their first interview on average 12 respectively 13 days after the start of fieldwork. On the other hand there are five countries in which it took more than five weeks until the interviewers realized their first interview. In Switzerland, Belgium, Luxembourg, and Ireland around 40 days were needed. In Germany even 80 days went by until the first interview was completed.

Figure 8 **Average number of days until interviewers completed their first interview in ESS Round 1**

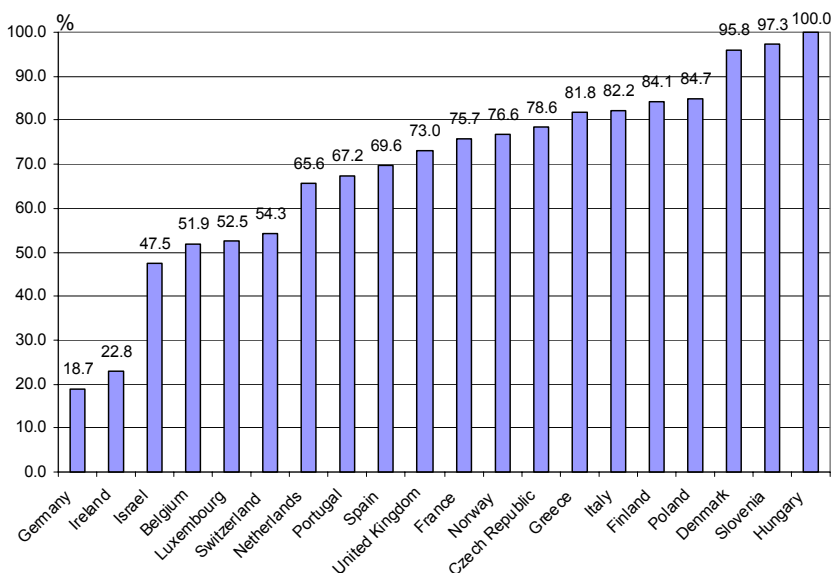


As a matter of course, these very long time spans until the completion of the first interview cannot be explained by the usual start-up phase of a survey. Unfortunately, it is not totally uncommon in survey practice that at the start of fieldwork for a survey not enough interviewers are available. Sometimes, a survey organisation does not have enough interviewers at all, or the interviewers are deployed at another study.⁷ So, it may happen, that some of the interviewers do receive the study materials and the addresses they have to work on not at the very beginning of fieldwork but sometime later on. In other cases, it can occur, that interviewers are available in theory, and all the sample addresses were allocated to them, but the interviewers decide by themselves to start later – for instance, because they put higher priority on a more attractive study from another survey organisation they are working for. Independent from the particular reasons which applied, it is certainly safe to say that according to our results for ESS Round 1 there are still opportunities to speed up the actual start of fieldwork in several countries.

In a different approach we can look at the proportion of interviewers who started their work within four weeks after the official start date of fieldwork (see Figure 9). In this approach outliers (i.e. very late starters) do not affect the results. The picture we get is very similar to the previous results. In three countries all or nearly all (95% or more) interviewers realized at least one interview during the first four weeks. These are Denmark, Hungary and Slovenia – the same three countries which also scored low on the average rate. In several countries only about half of the interviewers got started in the first four weeks: This is the case in Israel, Belgium, Luxembourg and Switzerland. Even lower is the proportion in Ireland and Germany. Here, only every fifth interviewer started to work successfully during the first four weeks of fieldwork.

7 We know, for instance, that the result for Germany is at least partly attributable to the deliberate decision to start fieldwork only with a few interviewers because the majority of interviewers was still working on a different project.

Figure 9 Percentage of interviewers with at least one completed interview within four weeks after start of fieldwork, ESS Round 1



4.4 Average length of fieldwork period on level of interviewers

Now, let us have a look at the working period of the interviewers, i.e. the time span between the date of the first and the date of the last interview of an interviewer. As a matter of course, it is possible that interviewers attempt to contact target persons before or after these dates. So, our indicator may somewhat underestimate the actual working period of the interviewers. Nevertheless our approach gives a good indication of the average interviewer working period, in particular, since we are mainly interested in comparing the results across countries.

Figure 10 **Average length of fieldwork period on level of interviewers in ESS Round 1 (in days)**

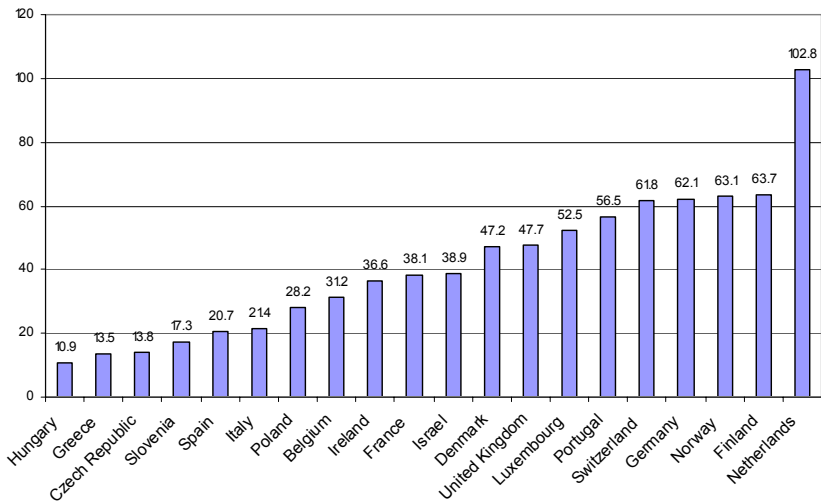


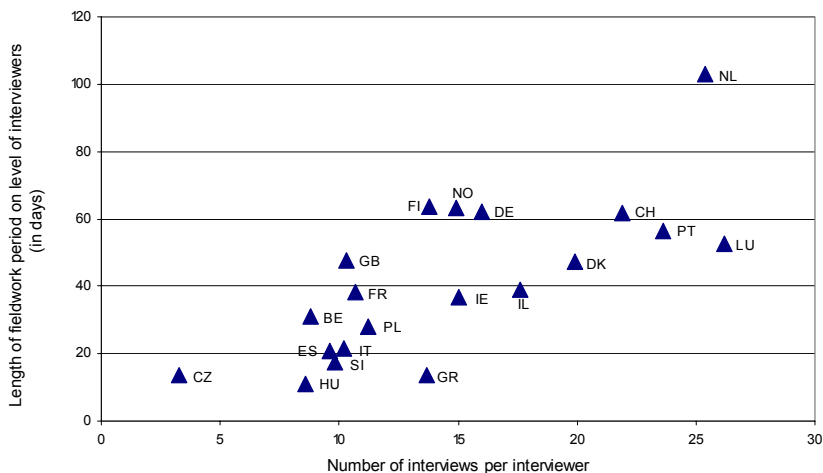
Figure 10 exhibits great differences in the length of the interviewers' working period across countries. In ESS Round 1 the average fieldwork period for the interviewers varies from a low of 11 days in Hungary to a high of 103 days in the Netherlands. The mean is 41 days across all countries.

As can be observed in Figure 11, the average number of days between the first and the last interview depends heavily on the average workload of the interviewers. The more interviews the interviewers had to make in a country, the longer the time period between first and last interview is ($r=.74$, $n=20$). But, still we find that even between countries with roughly the same average interviewer workload great differences do exist. In the two countries with the highest average interviewer workload for example – the Netherlands

and Luxembourg –, the fieldwork period of the interviewers is twice as high in the Netherlands as in Luxembourg. This outstanding high figure certainly has to do with the intensive (and succesfull) refusal conversion efforts which took place in the Netherlands (Philippens & Billiet, 2004).

On the other hand, there are two countries which exhibit a particular short effective fieldwork period compared to the other countries with equivalent interviewer workloads. These are Hungary (9 interviews in 11 days) and Greece (14 interviews in 14 days). One could doubt whether such a short average working period is long enough to allow for repeated contact attempts with people who are difficult to reach or difficult to persuade. But, in view of the high response rates which were achieved in both countries (70% and 80%, respectively), this did not seem to cause any real problems.

Figure 11 Average length of fieldwork period on level of interviewers (in days) and average number of completed interviews per interviewer, ESS Round 1



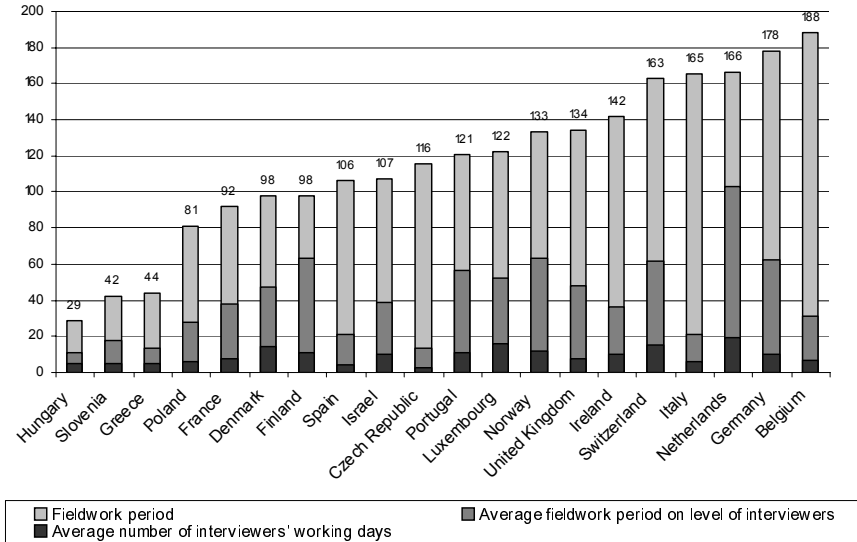
In general, the interviewers' working period is usually noticeably shorter than the whole fieldwork period in a country (see Figure 12). While the total fielding period varied between one to six months (29 to 188 days) across the 20 countries, the interviewers working period varied between 11 and 103 days on average. This means that knowing that a survey had a certain fieldwork period in a country must not be misinterpreted that all the interviewers were working for the survey all that time. In ESS Round 1 the average working period of the interviewers represents about 12%/13% (Czech Republic / Italy⁸) to 62%/65% (Netherlands / Finland) of the total fielding period⁹ For the average interviewer, the working period for the survey therefore was usually much shorter than the total fielding time.

This assertion holds even more, when we examine the average number of days at which the interviewers were taking interviews. The average number of (successful) working days for ESS Round 1 varies between three days in the Czech Republic and 19 days in the Netherlands. If we divide the average working period of the interviewers by the number of days at which the interviewers were completing interviews, we find that – across all countries – the interviewers were taking interviews every fourth day. The range goes from interviewing on average every second day in Hungary to interviewing on average every sixth day in Finland, Germany, and United Kingdom.

8 Italy seems to be a special case. Here, the fielding period comprised 165 days, but during week 10 to week 22 of the fieldwork period no interview at all was completed.

9 The higher the average number of interviews per interviewer, the higher the ratio of the average interviewers working period and the total fielding period ($r=.61$, $n=20$).

Figure 12 Length of fieldwork period, average length of fieldwork period on level of interviewers, and average number of interviewers' working days, ESS Round 1 (in days)



These results suggest the same conclusion as already drawn regarding the length of the start-up phase of interviewers: at least in principle, possibilities to reduce the fielding period seem to exist in several countries. The concrete possibilities will of course depend on the specific situation in a country. Countries, where the total fieldwork period was rather long in comparison to the average working period of the interviewers, could check for instance, whether some reduction in fielding time could be achieved by a more efficient, i.e. more simultaneous deployment of interviewers. Countries, where the number of working days of the interviewers was rather small in comparison to their average working period, could check whether the time span between the working days of the interviewers could be reduced. But, this recommendation does not mean that interviewers' working period should be reduced at any price. Of course, the interviewers still must have enough time to carry out all the procedures that are needed to obtain optimal response rates.

5 Correlates of Fieldwork Details – Two Examples

5.1 Fieldwork details and length of fieldwork period

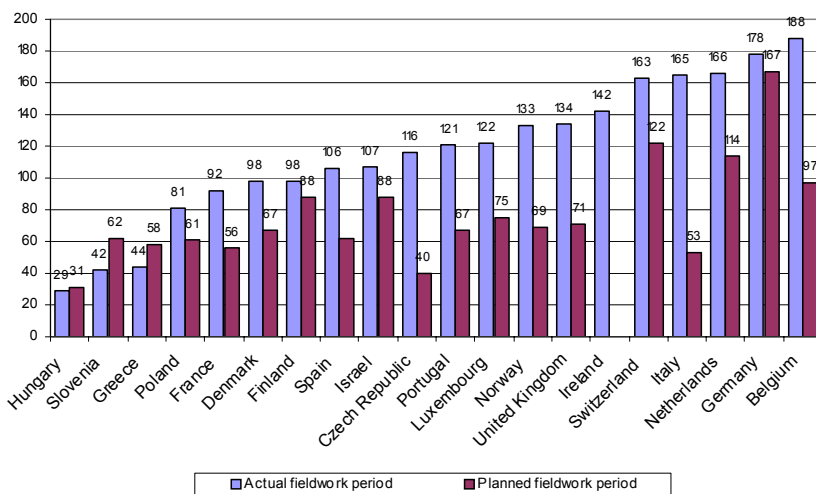
In cross-national surveys the parallel fielding of the survey in all participating countries is important, both for reasons of data comparability and for reasons of an effective management of the survey project. Since responses to survey questions can be affected by external events, it is prudent to strive for a fieldwork period being as parallel as possible in the participating countries. In ESS Round 1, for instance, the war in Iraq may have had a differential influence in different countries, depending on the timing of fieldwork (before, during and after the war). And with respect to the management of a cross-national survey, one has to be aware, that quite often a delay in data availability for one country can cause a delay of data delivery for the whole project, or it can require several data releases until all countries are included. In order to achieve simultaneous fielding periods in a multi-country survey, similar start dates and similar lengths of the data collection period in all participating countries are vital. In the present section we will deal with the latter issue in the context of ESS Round 1.

As a matter of course, a certain length of the fielding period cannot be legislated for. The length of the data collection period of a survey has to be planned and agreed upon with the survey organisation selected. The survey organisation has to define the necessary resources (personnel, laptops, etc.) and has to coordinate their availability with the needs of other studies the organisation is in charge of. The estimated length of the fieldwork period should be realistic, taking into account both the requirements of the given study (number of cases, availability and cooperativeness of targeted population, targeted response rate, etc.) and the resources available (number of interviewers to be deployed, their time budgets, available tools for fieldwork management, etc.).

In ESS Round 1, the Specifications laid down that the main fieldwork period should last at least one month within a four-month period in each country. A minimum length of one month was defined, in order to avoid that truncated fieldwork periods lead to a high proportion of non-contacts. A maximum of up to four months was thought to be sufficient to allow for repeated contact attempts at persons who are difficult to reach or difficult to persuade to participate.

As can be seen in Figure 13 nearly all countries deemed a fieldwork length between one and four months to be sufficient. The only exception was Germany with a planned length of 5,5 months.¹⁰ On average, the countries expected a fieldwork length of 76 days.

Figure 13 Actual and planned length of fieldwork period in ESS Round 1 (in days)*



*No information on length of planned fielding period available for Ireland

It turned out that the actual data collection overran its time. On average, fieldwork lasted 116 days in the 20 countries. There is no relationship between the expected length of fieldwork and the time span the fieldwork took longer than planned ($r = -.05$, $n = 19$). In two countries, Slovenia and Greece, the actual fieldwork period was three respectively two weeks shorter than planned. In Hungary, actual and planned length were virtually identical. In all the other countries fieldwork took longer than expected, by ten (Finland) to 112 days (Italy). As a result, the actual fieldwork length varied between 29 (Hungary) and 188 (Belgium) days across all countries.

10 The reason for that was, that Germany intended to start fieldwork at a particular time, when only a few interviewers were available because the majority of interviewers was still working on a different project.

In the following we will analyse whether these differences in the length of the fieldwork period can at least partly be explained by differences in the fieldwork patterns we described in the previous sections. A first hypothesis might be that the differences are related to differences in the sample size across countries. But, empirically we find only a weak correlation between the sample size and the length of the data collection period ($r=.18$, see Table 1). The relationship between the length of the fieldwork period and the number of interviewers deployed is even lower in size ($r=-.12$). Among the basic fieldwork parameters, the average number of realized interviews per interviewer shows the strongest correlation with the length of the fieldwork period ($r=.28$).

Table 1 **Correlation between length of fieldwork period and selected fieldwork patterns (Pearson's r , $n=20$ countries, ESS Round 1)**

	Pearson's r
<i>Basic information on fieldwork</i>	
Number of achieved interviews	.18
Number of interviewers	-.12
Average number of achieved interviews per interviewer	.28
<i>Fieldwork details</i>	
Average number of interviews per day per interviewer	-.51
Percentage of interviews made on weekday daytime	.39
Percentage of interviews made on weekday evening	.40
Percentage of interviews made on weekends	-.58
Average number of days until interviewers completed their first interview	.70
Average length of interviewers' working period (in days)	.55

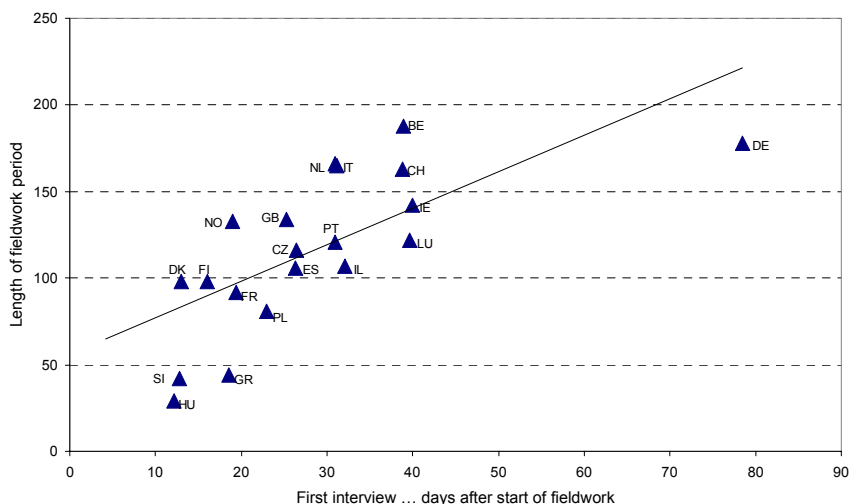
The situation looks somewhat different when we turn to the more detailed aspects of fieldwork. For all these variables, the correlation with the length of the fieldwork period is higher than for the basic fieldwork parameters. The higher the average number of interviews per interviewer per day in a country, the shorter is the fieldwork period in that

country ($r=-.51$, see Table 1). The same holds for the percentage of weekend interviews. Countries with a high percentage of weekend interviews exhibit shorter fieldwork periods than countries with a low percentage of weekend interviews ($r=-.58$). Both variables can be taken as indicators for the intensity of fieldwork: If interviewers complete a high number of interviews per day, and/or hold down their job not (only) on weekdays but (also) on weekends this can help to shorten the fieldwork period.

That a late start of fieldwork will lengthen the whole fieldwork period is in the nature of things. The correlation between the average start day of the interviewers and the length of the fieldwork period in all countries is $r=.70$. Also it comes by no surprise, that there is a relationship between the length of the interviewers' working period and the total length of the fieldwork period ($r=.55$). Interestingly enough, this correlation is smaller than the respective correlation between the average start date of the interviewers and the length of the fieldwork period.¹¹ That means, that for the total length of the fieldwork period it is more important when the interviewers actually start taking interviews than how long they are occupied with the study. In order to reduce the fielding time, it seems to be good advice to put a lot of emphasis on a timely start of all interviewers. Regressing the length of the fieldwork period on the average day of the first interview in ESS Round 1 results in an unstandardized regression coefficient of $b=2.11$ ($n=20$). That means that every single day the interviewers (on average) start later translates into a two days extension of the whole fieldwork period (see Figure 14).

11 When we put all four variables (number of interviews per day, percentage of weekend interviews, average start date and average working period of interviewers) into a multiple regression with the length of the fieldwork period as dependent variable, the average start day of the interviewers shows by far the strongest effect ($\beta=.59$, $n=20$).

Figure 14 Length of fieldwork period (in days) and average date of first interview (in days) in ESS Round 1



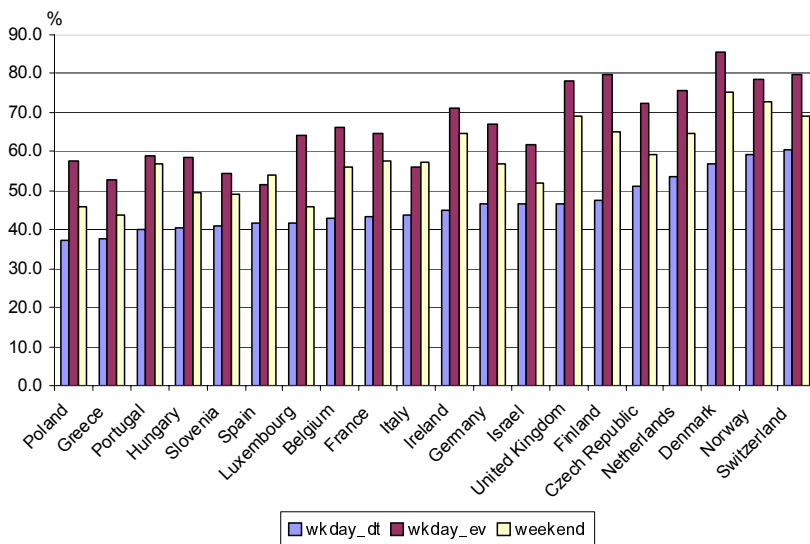
As a matter of course, we should not over-interpret these results. The analyses presented provide some useful hints on factors which can have an effect on fieldwork length, but the relationships we observed should not be taken for granted. First, and foremost we have to recognize that the data we analyzed do not stem from an experimental design, but are based on observations made on the aggregate level of countries. Second, it goes without saying that not only the variables analysed here will have affected the length of the data collection period in a country, but a lot of other factors as well. For instance, the population density and geographic distribution of the respondents, the cooperativeness of the target persons, or specific fieldwork efforts exercised by the survey organisations may have differed across countries, although we could not analyse their effects on the length of the fieldwork period in the present paper. This said we nevertheless believe that the results

we got shed some light on selected factors which may contribute to differences in the length of the fieldwork period across countries. Countries aiming at reducing their fieldwork period should check whether the factors analysed can be utilized to improve the implementation of data collection in future Rounds of ESS.

5.2 Timing of interviews and data quality

The timing of interviews can have an effect on data quality. Most textbooks on survey interviewing stress the fact that interviewers should do some of their work in the evenings or at weekends in order to interview people who are at work during the day (Groves et al., 2004: 293f; Hoinville et al., 1985: 107; Weinberg, 1983: 341f). When we use the data from ESS Round 1 and calculate the proportion of respondents who said that they had been in paid work during the last seven days separately for weekday daytime, weekday evening and weekend interviews, we observe strong and significant differences (see Figure 15).

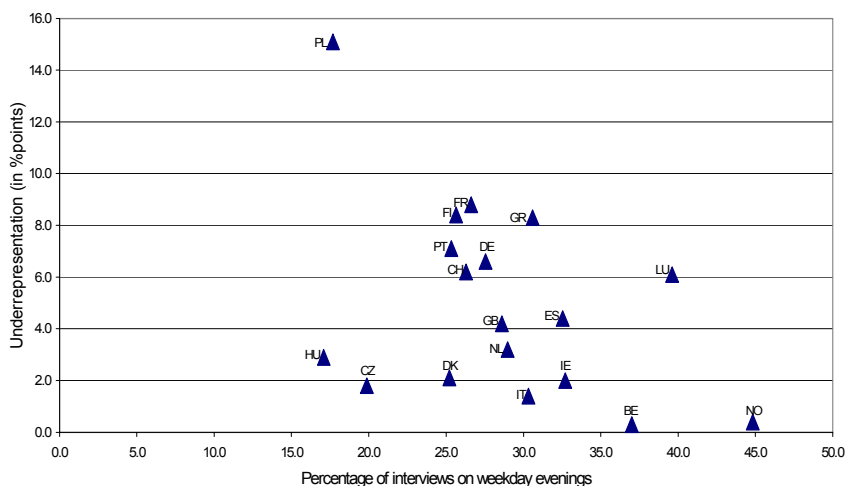
Figure 15 Percentage of people in paid work in the last 7 days, by time and day of interview, ESS Round 1 (data weighted by DWEIGHT)



In all countries, interviews made on weekday morning or weekday afternoon show the lowest proportion of people in paid work. In nearly every country, the highest proportion is observed for interviews made in the evening. The percentage of people in paid work is 10 (Spain) to 33 (Finland) percentage points higher for evening than for daytime interviews. Only in Italy and Spain, the proportion is slightly higher for weekend interviews than for evening interviews.

In order to validate the ESS data, we compared the labor force participation rates in the working-age population (15 to 64 years old) of the ESS with available external data. The external information stem from the “European System of Social Indicators” (EUSI, http://www.geis.org/en/social_monitoring/social_indicators/index.htm), which used data from Eurostat and the OECD, respectively (Indicator H1122). Information was available for 18 of the 20 ESS countries. No information is provided in EUSI for Israel and Slovenia. According to the EUSI, the labor force participation rates in the ESS countries vary between 60% (Italy and Hungary) and 82% (Switzerland). The proportions of people in paid work last week we receive from the ESS data are lower in all countries. But the differences vary considerably in magnitude across countries. In Belgium and Norway more or less negligible deviations of less than one percentage point were observed. On the other hand, we find differences of eight or nine percentage points for Greece, Finland, and France. The greatest deviation refers to Poland, where the labor force participation rate in ESS is 15 percentage points lower than the figure from EUSI.

Figure 16 Underrepresentation of people in paid work (in % points) and percentage of interviews completed on weekday evenings, ESS Round 1 (data weighted by DWEIGHT)



When we compare the magnitude of the deviations with the proportion of interviews realized on weekday daytime, on weekday evening or at weekends, we find only weak evidence for a relationship. The higher the percentage of interviews completed on weekday evenings, the lower the deviation with respect to the labor force participation rate ($r=-.43$, $n=18$; see Figure 16).¹² On the one hand, Poland shows the greatest observed deviation with respect to the proportion of people in paid work (15 percentage points), and the second lowest proportion of weekday evening interviews (18%).¹³ On the other hand, the lowest observed deviations pertain to two countries scoring well above average in the percentage of evening interviews: Belgium and Norway.

12 The respective correlations with the percentage of weekday daytime ($r=-.05$) and weekend interviews ($r=.27$) are lower in magnitude.

13 Given the great magnitude of this deviation, it seems worth to check whether other factors than nonresponse errors may have contributed to that bias, for instance differences in the translation of the respective question.

But besides that we have to be aware, that there are a few countries with a low percentage of evening interviews, which nevertheless do not show large deviations with regard to the labor force participation rate. A particularly striking example is Hungary, which exhibits a distribution of interviewing times very similar to that of Poland, but a deviation in the labor force participation rate of only three percentage points. And we have to take into account that the magnitude of the correlation we observed is heavily dependent on the data from Poland. Excluding Poland from the analyses reduces the correlation between the percentage of evening interviews and the underrepresentation of people in paid work to $r = -.24$ ($n = 17$).

These reservations make clear that a low percentage of interviews made in the evening does not inevitably lead to an underrepresentation of people in paid work in the realised sample. The main issue is, whether interviewers possess the necessary flexibility to complete an interview at a time which suits the preferences of the target person. According to our results, at least in some countries, a low proportion of evening interviews may reflect limited temporal availability of the interviewers, thus leading to an underrepresentation of people in paid work. Countries where this applies should check whether specific rules for the timing of interviews might help to improve the sample composition in future surveys.

For the rest, it is interesting to notice that another commonly used indicator for potential biases of survey results – the nonresponse rate – does not seem to be very indicative of an underrepresentation of people in paid work: In ESS 1, the correlation between the nonresponse rate and the underrepresentation of people in paid work is rather low in size, and it is in the other direction than one would expect ($r = -.20$, $n = 18$; excl. Poland $r = .03$, $n = 17$).

6 Conclusion

The present paper dealt with fieldwork issues in the first Round of the European Social Survey. Besides some basic features of data collection, like the number of interviews achieved or the number of interviewers deployed, several more specific aspects of fieldwork were investigated. These included the average number of interviews the interviewers completed per day, the time and day when interviews were made, the length of the start-up phase until interviewers completed their first interview, and the length of the time period interviewers were actually engaged in realizing interviews. The data presented

revealed differences and similarities in the work of interviewers across countries and provided by this means some insights in the freedom for different designs and implementation of fieldwork. We saw, for example, that

- the number of interviews completed per day was highest in South European countries, like Spain, Portugal or Greece,
- interviewing on weekends was a rather rare event in the Scandinavian countries,
- the time period until interviewers started to realize their first interview, and the length of the interviewers' actual working period differed widely across countries.

As it turned out, these fieldwork patterns can provide clues on how to explain the observed differences in the length of the fieldwork period or differences in sample composition across the ESS countries.

As a next step it could prove to be useful to examine whether the results we found can be replicated in other surveys – either in forthcoming Rounds of ESS or in other cross-national surveys (with the same countries but eventually with different survey organisations). In the present paper the three levels of 'country', 'survey organisation' and 'study' could not be distinguished, since they were perfectly confounded. Replications of the analyses carried out could help to clarify, whether the observed results pertain to stable patterns of fieldwork in particular countries and/or specific fieldwork organisations, or whether the results are more or less idiosyncratic features of the design and implementation of a specific survey.

The replication of the analyses is facilitated by the fact, that only a limited set of variables is needed. The list of variables we used comprises an interviewer identification number, as well as information on the date and start time of the interview. This information can be collected easily in practically every survey. Since we need this information only for respondents, the collection of the data can occur during the interview. Concentrating on information for realized interviews thus has the advantage of minimizing the effort necessary to gain the information. On the other hand, one has to bear in mind that more detailed analyses, for instance on the contact behavior of interviewers and its results, or nonresponse analyses investigating differences between respondents and nonrespondents, will not be feasible with these data. Such research questions require additional measuring instruments, like contact forms for instance (see Phillipens & Billiet, 2004 for an example of analyses based on contact forms data for ESS Round 1).

References

- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. Hoboken, New Jersey: Wiley.
- Hoinville, G., Jowell, R., & Associates. (1985). *Survey Research Practice*. Aldershot: Gower.
- Koch, A. (2002). 20 Jahre Feldarbeit im ALLBUS: Ein Blick in die Blackbox. *ZUMA-Nachrichten* 51, 9-37.
- Philippens, M., & Billiet, J. (2004, May). *Monitoring and evaluating nonresponse issues and fieldwork efforts in the European Social Survey*. Paper presented at the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen*. Opladen: Leske + Budrich.
- Weinberg, E. (1983). Data Collection: Planning and Management. In P. H. Rossi, J. D. Wright & A. B. Anderson (Eds.), *Handbook of Survey Research* (pp. 329-358). New York: Academic Press.

DISCUSSION OF THE INCOME MEASURE IN THE EUROPEAN SOCIAL SURVEY

A Proposal of Revised Survey Questions About the “Total Net Household Income”

UWE WARNER & JÜRGEN H.P. HOFFMEYER-ZLOTNIK

1 Introduction

Different types of surveys use different measurement instruments to assess “total net household income”. Depending on research questions and scientific purposes of the collected data, we distinguish three basic strategies to obtain income information about the surveyed household and its individual members:

- *Market research* is interested in categorizing the purchasing power of a household and classifies the households into consumer groups. The data producers focus on classes of income size and therefore they do not give a precise definition of income and they make no distinction between several surveyed population groups. They ask for the monthly amount of income and they give a general instruction in the question wording. The predetermined answer categories are income brackets. In case the interviewed person refuses to answer, the interviewer often takes the freedom to estimate the household’s income. An alternative used as economic status scale, formulated by ESOMAR (2003: 103), asks for “the penetration of ownership of ten key consumer durable items at household level”.
- *Social research* uses income as a socio-economic indicator on social stratification and inequality. In this approach the knowledge of size classes of the household net income is sufficient. In contrast to market research, social research surveys define the various income (main-) sources and formulate separate questions for different population groups. E.g. the wording of the income question differs for the self-employed and for employees. In Germany, the monthly net income is surveyed by an open question. To reduce non response a second question with income brackets is given to the interviewee in case of refusing the open question (Statistisches Bundesamt, 2004).

- *Economic and socio-economic research* is studying income distribution and the dynamics of changes in the economic situation of respondents and their households. The research question on the total income composition by various sources and the changes of the income types and amounts are of interest. Therefore a precise measurement of income is needed as well as for the entire household and its individual members. The several types of income are defined in detailed instructions to the interviewers and the interviewees and separated by their sources and types. Specific population groups and/or income recipients are interviewed according their socio-economic characteristics. For a defined income reference period (e.g. monthly) queries about gross and net income are asked through open questions; all possible monetary resources of all persons living in the household, as well as payments to the household per se are surveyed. In general the answer is given in gross and/or net amount (European Commission, 1996). In the European Community Household Panel (ECHP) income is measured by using a sixteen page long section in the person's questionnaire. Every member (fifteen years and over) of an eligible household answers the person questionnaire. The first step towards income information is a calendar on the respondent's labor force status in which month by month the employment situation is registered. The second step forward to income is a sequence of questions on having or not various income sources. After this the respondent is asked to give net and/or gross amounts of his/her income details during the income reference year. In addition one reference person is also surveyed by a household questionnaire. In this questionnaire five pages deal with incomes dedicated to and received by the household per se.

In this contribution we focus on social research. We are looking for an easy way to measure "total net household income" during social survey interviews in comparative perspective. "Easy" in this sense means a very limited number of questions, reasonable and practicable interview instructions, and a minimum interview burden for interviewee and interviewer. By decreasing the interview burden, we hope to increase the quality of the obtained income measure. "Comparable" here means that our target measurement quantifies in different countries and/or cultural contexts the same and comparable social fact. So far, as a "best practice" for cross-national/cross-cultural income questions we consider the European Social Survey.

We discuss the income measure in the European Social Survey and we propose revised questions to obtain the "total net household income" which will overcome the presented disadvantages of the used instrument.

2 Income Questions in the European Social Survey

In 2002, the first round of the European Social Survey (ESS) was carried out in 22 European countries (www.europeansocialsurvey.org).

One randomly selected household member was asked about the household income by the questions:

“Please consider the income of all household members and any income which may be received by the household as a whole. What is the **main** source of income in your household? Please use this card” (ESS 01/08/2002: F29). The show card 55 presented to the interviewee lists following main income sources:

- wages or salaries
- income from self-employment or farming
- pensions
- unemployment/redundancy benefit
- any other social benefits or grants
- income from investment, savings, insurance or property
- income from other sources

The question about the income amount is:

“... if you add up the income from *all sources*, which letter describes your *household's total net income*? If you don't know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly or annual income.” (ESS 01/08/2002: F30). The interviewee was confronted with show card 56 and should answer giving the corresponding letter.

CARD 56 YOUR HOUSEHOLD INCOME

	Approximate WEEKLY	Approximate MONTHLY	Approximate ANNUAL	
J	Less than €40	Less than € 150	Less than €1800	J
R	€40 to under €70	€150 to under €300	€1800 to under €3600	R
C	€70 to under €120	€300 to under €500	€3600 to under €6000	C
M	€120 to under €230	€500 to under €1000	€6000 to under €12000	M
F	€230 to under €350	€1000 to under €1500	€12000 to under €18000	F
S	€350 to under €460	€1500 to under €2000	€18000 to under €24000	S
K	€460 to under €580	€2000 to under €2500	€24000 to under €30000	K
P	€580 to under €690	€2500 to under €3000	€30000 to under €36000	P
D	€690 to under €1150	€3000 to under €5000	€36000 to under €60000	D
H	€1150 to under €1730	€5000 to under €7500	€60000 to under €90000	H
U	€1730 to under €2310	€7500 to under €10000	€90000 to under €120000	U
N	€2310 or more	€10000 or more	€120000 or more	N

(Source: ESS 01/08/2002: Card56)

3 Factors Having an Impact on the Measurement Quality

In previous research we analyzed which five factors having an impact on the measurement quality by ESS questions. These are:

1. *The impact of household definitions:*

Nearly all European countries use different definitions of household and different household definitions have effects on the household size – more precise: who is counted as household member. The blue print of ESS questionnaire is using the definition of households applicable in England: “One person living alone or a group of people living at the same address (and have that address as their only or main residence), who either share at least one main meal a day or share the living accommodation (or both).” (ESS 15/07/2002: 11). This statement is made in the Project Instructions meant for the interviewers; no definition is given to the respondent during the interview. So each respondent has the freedom to apply commonly used definitions of household. For illustration, only a small number of household definitions are listed:

- In Germany, the household definition focuses on the common kitchen. Therefore an apartment sharing community is defined as one household if the occupants are cooking, preparing meals together. If each of the people of this community will cook for his or her own, then in the same apartment splits into each person’s own household.
- In Italy, the household is defined by the common atrium. By this definition one household may occupy more than one dwelling.
In addition, the Italian part of ESS uses “family” in the question wording during the interviews: “totali nette della sua famiglia”. (ESS 2002, VERSIONE ITALIANA: 19-12-02: F30) It is obvious that “family” constitutes a different membership then household definition does.
- In Luxembourg, the shared living room identifies the household unit.

2. *The impact of the main income source:*

If main income source is related to income from work or former labor market activities, like wages and salaries or pensions, then main source covers a big amount of the total net household income. This (one) amount can be calculated and remembered by the interviewee. An increasing number of income sources, others than income from work, increases the complexity of adding up the household income sum. In this case of income composition by sometimes small monetary quantities instead of one main

source respondent can be disoriented. For households with arrangements of social transfers as the main source, often the total household income is underreported (Warner & Hoffmeyer-Zlotnik, 2005: 213-215).

3. *The impact of income composition:*

“Total net household income” is a composition of many different gains per household. In the ECHP questionnaire where income is asked source by source for each individual, there person's income consists of up to ten different sources. Across the countries involved in ECHP the mode is between 5 and 7 different sources per person (Warner & Hoffmeyer-Zlotnik, 2005: 215-216).

4. *The impact of the respondent's family relation to the main income earner:*

The ESS sample design selects randomly one household member as interview partner. A responding person can have a close family relationship to the main income earner. These are main income earner him- or herself and his or her partner. The other cases like the children and/or the parents and/or other relatives we interpret as interviewees, having a distant relation to the main bread winner of the household. A comparison of the interview outcomes of ESS and ECHP shows that the closer the relation to the main income earner the better the information about the monetary situation of the household is (Warner & Hoffmeyer-Zlotnik, 2005: 212-213).

5. *The impact of remembering income:*

Having in mind the income composition of a household with mostly five or seven different sources and the respondent's position inside the observed household, it is obvious that the interviewee can be good informed and also can be less informed about the socioeconomic situation of the entire household. Analyzing the ESS income answers by the grade of information, we noticed that in most countries less informed persons dominate the low income categories (Warner & Hoffmeyer-Zlotnik, 2005: 216-218).

4 Categorizing Income for Comparative Social Research

In all national surveys, the European Social Survey uses the same twelve income brackets as answer categories. Are these income brackets optimal answer categories for poor as well for rich populations? By cutting the “total net household income” variable of ECHP wave 8 into 5% groups of the population and sorting the ESS categories into this attained distribution, we illustrate the need to adjust the income brackets to national financial circumstances and the national income distributions (see Table 1 for selected countries).

Table 1 The distribution of the nineteen 5% percentiles from ECHP8 by the 12 income categories of ESS in selected countries

ESS categories	Germany	United Kingdom	Italy	Luxembourg	Portugal	Finland
	No. of the ECHP (wave 8) 5%-percentile					
up to 1,800	---	---	---	---	---	---
1,800 – 3,600	---	---	---	---	1-2	---
3,600 – 6,000	---	---	1	---	3-5	---
6,000 – 12,000	1-2	1-2	2-5	---	6-11	1-3
12,000 – 18,000	3-5	3-5	6-10	1	12-15	4-7
18,000 – 24,000	6-8	6-7	11-13	2-3	16-17	8-10
24,000 – 30,000	9-12	8-10	14-16	4-6	18	11-12
30,000 – 36,000	13-14	11-12	17	7-8	19	13-15
36,000 – 60,000	15-19	13-17	18-19	9-15	---	16-19
60,000 – 90,000	---	18-19	---	16-18	---	---
90,000 – 120,000	---	---	---	19	---	---
120,000 and more	---	---	---	---	---	---

Source: ECHP UDB version April 2004, own calculations

The ESS category “36,000 € to 60,000 €” covers the 9th to the 15th 5% percentiles of the income distribution in Luxembourg. In Germany, the same income category covers the 15th to 19th 5% percentiles. In Portugal, the richest 5% of the population have a total net household income of “36,000 € to 60,000 €”. Also, the poorest 5% of the Luxembourg people have a higher household income than 75% of the Portuguese population and 50% of the Italians.

Respondents from all countries need about six ESS categories to answer the income question. But they use different six categories in different types of countries.

5 Proposal for Revised Questions and Answer Categories

With respect to the factors having an impact on the measurement quality of our target variable, we reformulate and specify the income questions of the ESS. The income questions in ESS started by asking about the main income source of the household. But in a household there are several persons with an average of 5 to 6 different income sources each. Therefore the starting question should generate a feeling about the whole range of possible income sources. A list of income sources should help respondent remembering the sources and later on summing-up the amounts.

Income question 1:

“Please consider the income of every member of the household and any income which may be received by the household as a whole. What are the sources of income in your household? Please tick all applicable.”

The show card 1 facilitates the respondent's task during the interview and the listed income types cover the whole range of income sources. The categories are described by keywords and groups of explaining terms (see Table 2). Only superordinate monetary concepts are used because income sources are determined by national tax- and welfare-system. Therefore in each country different numbers of income components are in effect for each income title (see: Canberra Group, 2001: 170-177).

Table 2 **Categories on show card 1, for income question 1**

ALL INCOME SOURCES OF YOUR HOUSEHOLD
Employee income, including bonuses (e.g. vacation or Christmas), tips, extra payments (from e.g. overtime and shift work), profit sharing
Income from self-employment or farming, also free-lance work
Pensions, including old age and widow's pensions, retirement
Unemployment / redundancy benefits, including benefits related to training and sickness allowances
Rentals and Property income
Current public transfers received, social benefits and grants including child and family allowances, universal and/or means-tested social assistance and orphan's pensions, educational grants
Regular private transfers from persons outside your own household including alimony
Income from other sources including reimbursements from taxes and insurances, lottery winnings

Question 2 is the core income question. The respondent is asked to add-up the amount from all sources for all household members to get the "total net household income". A specific problem of income measurement is the definition of net income. Therefore the instruction to the respondent comprise in the question not only in the project instructions for the interviewer the definition and calculation of net income.¹

Question 2:

"If you add up the income from all sources and all household members (from the target population), which letter describes your household's total net income? Net is after

1 The definition of household should be the same for all parts where information about household were asked. Therefore the definition of household is not a specific problem of income measurement.

deduction of national taxes and after deduction of compulsory contributions to the national social security. If you don't know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly or annual income."

We also allow an estimation if the interview partner can not reconstruct the exact figure. In the interviewer's instructions we demand for the information whether respondent has calculated or estimated the amount. This allows us to flag the answer quality.

Table 3 Show card 2.1, question 2: Proposed categories for low income countries; tested for Portugal and Italy

YOUR NET HOUSEHOLD INCOME				
	Approximate WEEKLY	Approximate MONTHLY	Approximate ANNUAL	
M			Less than 2,500€	M
B			2,500 to under 5,000€	B
F			5,000 to under 7,500€	F
G			7,500 to under 10,000€	G
Q			10,000 to under 12,500€	Q
N			12,500 to under 15,000€	N
T			15,000 to under 20,000€	T
D			20,000 to under 25,000€	D
K			25,000 to under 30,000€	K
W			30,000 to under 35,000€	W
H			35,000 to under 40,000€	H
C			40,000 to under 45,000€	C
J			45,000 to under 50,000€	J
U			50,000 to under 55,000€	U
I			55,000 to under 60,000€	I
Z			60,000€ and more	Z

For illustrative purpose we filled the annual amount only

The show card presenting the income categories to the respondent varies across the countries and the empirically found income distribution in the observed societies. For the EU-15 countries, we propose at least three different show cards with income brackets necessary: one show card for nations with a low average income like in Portugal or Italy, one show card for nations with middle average income like in United Kingdom, in Germany, or in Finland, and one show card for nations with high average income like in Luxembourg (see Tables 3 to 5). Because of data availability we have not tested countries like Bulgaria. We assume, a fourth show card becomes indispensable for nations with very low income distributions.

For countries with low average income the categories about “annual” income start at “less than 2,500 Euros”. In this type of countries slow steps upwards the income distributions are needed. Therefore the next five categories increase by steps of 2,500 Euros. At category seven we continue in 5,000 Euro steps. The highest income group is the 16th category of about “60,000 Euros and more”. For the column “monthly”, the categorical scheme starts in steps of 200 Euros, continuing with the double amount (400 Euros) at the 7th category.

Table 4 Show card 2.2, question 2: Proposed categories for middle income countries; tested for United Kingdom, Germany and Finland

YOUR NET HOUSEHOLD INCOME				
	Approximate WEEKLY	Approximate MONTHLY	Approximate ANNUAL	
O			Less than 5,000€	O
V			5,000 to under 10,000€	V
L			10,000 to under 15,000€	L
T			15,000 to under 20,000€	T
D			20,000 to under 25,000€	D
K			25,000 to under 30,000€	K
W			30,000 to under 35,000€	W
H			35,000 to under 40,000€	H
C			40,000 to under 45,000€	C
J			45,000 to under 50,000€	J
U			50,000 to under 55,000€	U
I			55,000 to under 60,000€	I
S			60,000 to under 70,000€	S
E			70,000€ and more	E

The categories for countries with middle average income we start at “less than 5,000 Euros annually”. In this type of countries steps of 5,000 Euros are adequate. Therefore nearly all of the 14 categories are increasing in steps of 5,000 Euros. Only the second last step is about 10,000 Euros. The last, the 14th category is about “70,000 Euros and more”. If the “monthly” amount is preferred by the interviewee the categorical scheme proceeds in steps of 400 Euros.

Table 5 Show card 2.3, question 2: Proposed categories for high income countries; tested for Luxembourg

YOUR NET HOUSEHOLD INCOME				
	Approximate WEEKLY	Approximate MONTHLY	Approximate ANNUAL	
O			Less than 10,000€	O
L			10,000 to under 15,000€	L
T			15,000 to under 20,000€	T
D			20,000 to under 25,000€	D
K			25,000 to under 30,000€	K
W			30,000 to under 35,000€	W
H			35,000 to under 40,000€	H
C			40,000 to under 45,000€	C
J			45,000 to under 50,000€	J
U			50,000 to under 55,000€	U
I			55,000 to under 60,000€	I
S			60,000 to under 70,000€	S
Y			70,000 to under 80,000€	Y
X			80,000 to under 90,000€	X
A			90,000 to under 100,000€	A
R			100,000 to under 110,000€	R
P			110,000 € and more	P

In countries with a high average income, like Luxembourg, the average income of the poorest 5 % of the population is higher than the average income of the poorest 50 % of Portugal's population. In countries with a relatively high average income we begin the "annual" categories at 10,000 Euros. The following steps increase by 5,000 Euros. Reaching the benchmark of 60,000 Euros the categorical ladder continues with ranges of 10,000 Euros. The highest (the 17th) income category is "110,000 Euros and more". If the "monthly" amount is preferred by the respondent the given income scheme starts in steps of 800 Euros.

A common cross country coding is possible for our income schemes: The starting step for low average income is 2,500 Euros, for middle average income distributions it is 5,000 Euros, and for high average income countries we begin at 10,000 Euros. From the category 15,000 Euros on all three schemes are in the same rhythm. All three categorical schemes end at different top income brackets, but starting at 60,000 Euros the increment is 10,000 Euros, in each country as often as needed to reach the higher end of the income distribution.

Question 3 offers the possibility to put household income in relation to the total number of income earner. We propose to ask for the number of household members contributing to the household's total net income:

Question 3:

“How many household members contribute to the household's total net income?”

In question 4 we follow the original ESS questionnaire and survey the main source of household income. Differing from ESS questionnaire we are use again show card number 1 a second time (as show card 3). Show card 1 contains the categories covering the whole range of income sources and is comparable to those categories the respondent used short time ago in question 1. But this time only one answer is possible, the main source of income.

Question 4:

“Please consider the income of every member of the household (from the target population) and any income which may be received by the household as a whole. What is the main source of income in your household? Only one answer possible.”

In our last question we evaluate the validity of the answers, so far obtained. One of our hypotheses was: the closer the relation to the main income earner the better the information about the monetary situation of the household. For evaluate the answer quality the data user needs information about the closeness of respondents, selected for the interview inside the sampled household, to the main bread winner. Therefore we ask:

Question 5:

“Who is the main income earner of your household?”

The categories on show card 4 to become aware of the main income earner are:

- myself
- my partner/spouse
- myself and my partner/spouse
- my father and/or my mother
- my child
- other member of the household

6 Conclusion

We have shown that the used measurement instrument of the European Social Survey can be optimized. Income measurement in social surveys is a complex task for all actors in interview situations. On the one hand, a lot of impact to the answers quality is provoked by the respondent. On the other hand, the categories of show cards concerning the income amounts and types are influenced by the national tax regulation and the national income inequality. In the case of ESS, we assume an impact on the income information by the fact that the predetermined income brackets are dominated by the country where the blue print of the harmonized questionnaire is originated. We recommend strongly testing the income questions across countries and for various income distributions to make sure that the surveys measures in all participating countries the actual monetary inequality of households.

With less interview burden we obtain information appropriate for sociological research. Our revised instrument offers the requirements to measure income detailed enough. The major characteristics having an impact on the answer quality are controlled during the interview situation. Our offered system of answer categories consists of three different types of categorical systems and reflects the national income distribution and is at the same time coordinated across countries. By using 5 questions, the results from comparative surveys become meaningful and significant for the interpretation in social sciences.

References

- ESOMAR (2003). The ESOMAR Standard Demographic Classification. A System of International Socio-Economic Classification of Respondents to Survey Research. In Hoffmeyer-Zlotnik, J. H.P. & C. Wolf (eds.). *Advances in Cross-National Comparison. A European Working-Book for Demographic and Socio-Economic Variables* (Pp 97-121). New York: Kluwer Academic/Plenum Publisher.
- European Commission (1996). *European Community Household Panel (ECHP): Volume 1. Survey Methodology and Implementation*. Luxembourg: Survey Questionnaires.
- European Commission/Eurostat (2003). *ECHP UDB Description of Variables. Data Dictionary, Codebook and Differences between Countries and Waves*. (DOC.PAN 166/2003-12). Luxembourg.
- European Commission/Eurostat (2003). *ECHP UDB Construction of variables. From ECHP questions to UDB variables*. (DOC.PAN 167/2003-12). Luxembourg.
- European Social Survey (2002). *The European Social Survey*. Source Questionnaire (Round 1, 2002). ESS Document Date. 01-08-02. <http://www.europeansocialsurvey.org/>

- European Social Survey (2002). *The European Social Survey*. Source Questionnaire (Round 1, 2002). TNS ABACUS – Versione Italiana: 19-12-02. <http://www.europeansocialsurvey.org/>
- European Social Survey (2002) *The European Social Survey*. Source Showcards. <http://www.europeansocialsurvey.org/>
- European Social Survey (2002) *The European Social Survey*. Project Instructions (PAPI). ESS Document Date. 15-07-02. <http://www.europeansocialsurvey.org/>
- Expert Group on Household Income Statistics, The Canberra Group (2001). *Final Report and Recommendations*. Ottawa: The Canberra Group
- Statistisches Bundesamt (ed.) (2004). *Demographische Standards*. Ausgabe 2004. <http://www.gesis.org/Methodenberatung/Untersuchungsplanung/Standarddemografie/index.htm>
- Warner, U. & Hoffmeyer-Zlotnik, J. H.P. (2005). Measuring Income in Comparative Social Research. In Hoffmeyer-Zlotnik, J. H.P. & J. A. Harkness (eds.), *Methodological Aspects in Cross-National Research* (Pp 203-221) ZUMA-Nachrichten Spezial 11. Mannheim: ZUMA.

AN EXAMINATION OF PRETESTING METHODS FOR MULTICULTURAL, MULTILINGUAL SURVEYS

The Use of Cognitive Interviews to Test Spanish Instruments

PATRICIA L. GOERMAN

1 Introduction

Pretesting methods that are currently used in survey research to develop and improve survey questions have been evolving for the past 20 years. A widely accepted starting point for the formalization of pretesting techniques, which were originally adapted from cognitive psychology, is the Cognitive Aspects of Survey Methodology (CASM) Conference (Jabine et al., 1984). These early efforts to develop formal pretesting methods were later followed by the formal implementation of pretesting techniques at many federal statistical agencies (e.g., U.S. Census Bureau, 2003). Since then, a number of studies have been implemented to assess the relative merits of different pretesting methods (e.g., Presser, et al., 2004).

While there has been increased focus on issues related to multicultural and multilingual survey design in recent years, (e.g. Harkness, Van de Vijver & Mohler, 2004) there has been relatively little research on the appropriateness of specific pretesting methods and techniques with respondents from different cultural and/or linguistic groups. At the same time, many researchers have identified difficulties in using widely accepted cognitive interview techniques with different cultural and linguistic groups (for example, Pan, 2004; Coronado & Earle, 2002; Blumberg & Goerman, 2000; Kissam, et al., 1993).

This paper lays out a research plan designed to continue the process of tailoring and refining cognitive interview techniques to enhance their effectiveness across cultural and linguistic groups, with a focus on Spanish-language interviews in particular. It includes 1) a review of the literature on pretesting multilingual surveys, and 2) plans for an exploratory pilot study designed both to identify difficulties in conducting cognitive interviews with Spanish-speaking respondents and to identify and test alternative methods. The paper concludes with a brief discussion of research on pretesting Spanish instruments that is currently underway at the U.S. Census Bureau.

2 Review of the Literature on Pretesting Multilingual Surveys

2.1 The increasing need for multilingual survey instruments

Because of growth in globalization and migration in recent decades, there has been an increasing need to conduct surveys in multiple languages in order to ensure that data are representative of all members of culturally and linguistically diverse populations. This has been the case for both research that crosses national boundaries and research within single countries. Within the United States alone, the Census Bureau identified 380 categories of individual languages or language groups in 2000 (Shin & Bruno, 2003).

There are many people in the United States who have difficulty communicating in English. The Census Bureau reported that 4.4 million households, encompassing 11.9 million people, were linguistically isolated in 2000.

The Census Bureau defines a linguistically isolated household as "...one in which no person aged 14 or over speaks English at least 'very well.'" A linguistically isolated person is defined as "...any person living in a linguistically isolated household." It should be noted that this definition can include household members under 14 years of age who do in fact speak English fluently or very well (Shin & Bruno, 2003). The numbers of linguistically isolated households and people are up dramatically from the 1990 Census, in which 2.9 million households and 7.7 million people were classified as linguistically isolated.

In the United States, the proportion of people over age five who spoke a language other than English at home grew from 14% or 31.8 million people in 1990 to 18% or 47 million people in 2000 (Shin & Bruno, 2003). Spanish was the most common non-English language spoken in the home in 2000, with 28.1 million speakers. Almost half of these people reported that they spoke English less than "very well." Chinese was the second most common language spoken in the home with 2 million speakers, followed by French, German, Tagalog and Vietnamese (Shin & Bruno, 2003).

Both federal and private survey research organizations throughout the world are becoming increasingly aware of the potential bias inherent in the use of predominantly monolingual and monocultural survey instruments to measure diverse populations (Harkness, et al., 2004; McNally, 2000; Rogler, 1999). In recent years, the U.S. Census Bureau has made many strides towards providing multilingual survey instruments. For example, the 2000 Census of Population and Housing was available to respondents in five non-English languages: Spanish, Chinese, Korean, Vietnamese, and Tagalog. In addition, language guides were available in 49 languages and were used to assist respondents in filling out their answers on English language questionnaires. Previously, the 1990 Census had provided

paper questionnaires in English and Spanish and language assistance guides in 32 languages (Briggs, 2001). The Census Bureau has recently issued both a Translation Guide-line and a Pretesting Standard, which will greatly improve the quality of newly created survey translations (U.S. Census Bureau, 2003, 2004).

2.2 Pretesting multilingual instruments

An examination of current practices related to multilingual surveys shows that, unfortunately, it is not common practice to pretest all language versions of a survey before it is finalized (Harkness, 2004; Willis, 2004; Potaka & Cochrane, 2004). Survey designers often do extensive pretesting of the original version of a survey, but there are a number of barriers to thorough pretesting of translations. First of all, pretesting is time-consuming and can require a great deal of financial resources. Pretesting multiple versions of a survey increases this burden. In addition, there is pressure for researchers to maintain question wording that has been used in the past. Historically, there has been a belief that translations should be kept as similar as possible to the original (Harkness, 2004). There are additional barriers to the extensive pretesting of multilingual surveys, such as difficulty in hiring the appropriate bilingual or multilingual pretesting staff, and difficulty recruiting respondents from various cultural and language groups (Willis, 2004).

Despite these challenges, there has been an increase in the number of multilingual instruments being pretested in recent years (e.g. Pan, 2004; Carrasco, 2003; Potaka & Cochrane, 2002, 2004; Coronado & Earle, 2002; Blumberg & Goerman, 2000; Kissam, et al., 1993). Many researchers have identified limitations in the use of widely accepted cognitive interview techniques when used with non-English-speaking populations in the United States.

2.3 Common cognitive interview methods

Widely accepted cognitive interview techniques include a variety of methods (Willis, 2005; Gerber, 2004; Pan, 2004; Sirken, et al., 1999). For example, many researchers use think-aloud protocols, for which respondents are asked to “think out loud” as they choose responses to survey questions, or fill out a paper questionnaire. In addition, there are a number of widely used cognitive interview probes. First of all, “meaning-oriented probes” are commonly used to ask about a respondent’s interpretation of a term, phrase or entire question. A typical meaning-oriented probe would be, “What does the term *foster child* mean to you in this question?”

“Process-oriented probes” are those which ask a respondent to explain “the process by which [he or she] calculates an answer, decides between alternative answer categories, or makes a judgment about an answer” (Pan, 2004). A typical process-oriented probe would

read “How did you arrive at/choose that answer?” “Recall probes” are often used to determine ways in which a respondent recalls information; for example, one might ask, “How do you remember that *you’ve been living in your house for 17 years?*” A final example of a common cognitive interview technique is that of paraphrasing. In order to find out whether a respondent is interpreting a question in the way that survey designers intended, a “paraphrasing probe” would ask the respondent, “Can you tell me in your own words what this question is asking?”

2.4 Cognitive interview challenges

It is important to keep in mind that low educational level, English-speaking respondents in the U.S. have been found to have difficulty with cognitive interview techniques in general (Willis, 2005; Miller, 2003; Bickert & Felcher, 1996; Wellens, 1994). For example, Willis has found paraphrasing to be difficult for respondents with low educational levels. Bickart & Felcher and Wellens have found that low educational level respondents often have great difficulty in producing think-aloud protocols. However, it should be noted that Wellens’ respondents with limited English proficiency were interviewed only in English.

Finally, a lack of experience with the survey interview context in general and difficulty with certain types of survey questions in particular have been shown to cause English-speaking respondents of low educational levels to experience great discomfort in the cognitive interview setting (Miller, 2003). Within the United States, many linguistically isolated people tend to have low levels of education, so in this context one must take care not to confound these two variables.

2.5 Cognitive interviews in non-English languages in the U.S.: Chinese speakers

Pan (2004) has found that a number of widely accepted cognitive interview techniques are problematic when adapted for use with Chinese speakers living in the U.S., regardless of their educational level. Pan conducted 10 cognitive interviews with Chinese-speaking respondents based on the Chinese paper and pencil version of the U.S. Decennial Census questionnaire. She found that regardless of their educational backgrounds, most of her immigrant respondents were not familiar with the social context of a survey interview. In many cases this caused them to interpret the task at hand as a sort of test, where they needed to provide the correct answer as opposed to their own interpretations or opinions of question wording. Despite her repeated attempts to explain the task at hand, respondents asked for constant reassurance and often exhibited defensiveness during the interview process.

Pan also found that standard cognitive interview probes were difficult to translate into Chinese and often cultural equivalents for the terms and concepts in question simply did not exist. She had difficulty getting respondents to understand and respond to both think-aloud instructions and process-oriented probes. Pan found that meaning-oriented probes were generally more effective with Chinese-speaking respondents. This type of probe was most effective when she asked respondents to describe their interpretation of specific words or phrases, but was often ineffective when she asked them about their interpretation of entire questions.

Most notably, Pan's respondents experienced these difficulties with cognitive interview techniques regardless of their educational level. She attributes many of their difficulties to differences in cultural frames of reference, differential experience with the survey process, and problems with the translation of the terms and concepts involved in typical cognitive interview probes. It remains to be seen whether and to what extent these findings will apply to other cultural and linguistic groups, such as Spanish-speaking immigrants in the United States, the focus of the present research.

2.6 Cognitive interviews in non-English languages in the U.S.: Spanish speakers

Researchers conducting cognitive interviews with Spanish-speaking respondents in the U.S. have also encountered difficulties in applying standard cognitive interview techniques to their respondents (e.g. Blumberg & Goerman, 2000; Coronado & Earle, 2002; Kissam, et al., 1993). None of these studies was specifically designed to evaluate cognitive interview techniques, but all mention difficulties encountered by the researchers.

Similar to Pan's findings with Chinese-speaking respondents, Coronado and Earle noted a great deal of difficulty in clarifying to Spanish-speaking respondents that the cognitive interview itself was not a "test." Many of their respondents expressed a lack of confidence and needed frequent reassurance that they were not answering incorrectly and even that they were not the "wrong" person to be serving as a respondent. Kissam, et al. (1993) similarly found that many of their Spanish-speaking respondents interpreted the cognitive interview as a "test" and that they expressed "anger, disgust or humiliation" at their own performance. A number of their respondents even asked that the tape recorder be turned off periodically, so as not to record their requests for clarification or assistance.

Coronado and Earle found that Spanish-speaking respondents had difficulty with meaning-oriented probes and most often repeated questions verbatim when asked to paraphrase them. In addition, people often became irritated with what they perceived to be the repetitive nature of the probe questions. Respondents also provided a great deal of extraneous commentary before getting to what the interviewers perceived to be the "point" of the questions.

In cognitive interviews of mainly undocumented Spanish-speaking immigrants in the U.S. for a Census Bureau study, I found that I had to spend inordinate amounts of time explaining the purpose of the interview and reassuring people about confidentiality (Blumberg and Goerman, 2000). In addition, the standard consent forms and payment vouchers used by the Census Bureau for the conduct of cognitive interviews were major stumbling blocks for some respondents, both in terms of difficulty reading them and in terms of respondents' fears about confidentiality. On the whole, a great deal of reassurance and extraneous conversation was necessary to put respondents at ease.

While increasing numbers of multicultural and multilingual surveys are being conducted and pretested, there remains a great deal of uncertainty as to the best practices and methods for pretesting different language versions of multilingual survey instruments. This paper outlines a research plan designed to examine the effectiveness of widely accepted cognitive interview techniques with Spanish-speaking respondents in the U.S.

3 Exploratory Research Plan

My exploratory study will have three phases and it represents the initial step towards the ultimate goal of building on our existing knowledge of how cognitive interview techniques can be used to develop and improve data-collection instruments. In particular, my research will focus on ways in which pretesting techniques can best be modified and adapted for use with non-English language data-collection instruments.

3.1 Research Phase 1

As a part of phase 1 of the research, I will conduct exploratory cognitive interviews with 20 monolingual, Spanish-speaking respondents of a variety of national origins or backgrounds (see Table 1). Ten of these respondents will be of lower educational levels, having less than a high school diploma or 12 years of schooling. The other ten respondents will have a high school level diploma or higher. I will conduct the interviews using classic think-aloud techniques and both concurrent and retrospective probing methods. Concurrent probes are questions asked throughout the course of a cognitive interview, immediately following each survey question. Retrospective probes are often asked as part of a debriefing section at the end of the cognitive interview. Many researchers use a combination of concurrent and retrospective probing within the same interview, for example, asking a block of survey questions followed by a series of probes.

I will create a cognitive interview protocol by translating common English language cognitive interview instructions and probes into Spanish.

Table 1 Summary of respondent characteristics for Phase 1 of the research

	Spanish speakers
Low educational level (-12 years)	10
High educational level (12+ years)	10
Total respondents	20

3.2 Research questions

The goal of this study will be to systematically identify some of the difficulties experienced by these respondents. Examples of my research questions are listed below.

1. How can the cognitive interview task best be introduced to Spanish-speaking respondents to avoid the perception of a “test” situation?
2. Are standard cognitive interview probes appropriate for Spanish-speaking respondents; do they make sense to them?
3. How are standard cognitive interview probes interpreted by Spanish-speaking respondents?
4. How can the goals of standard probes be better achieved for this population?

These interviews will be based on a segment of the CAPI (Computer Assisted Personal Interviewing) Spanish version of the U.S. Census Bureau’s American Community Survey and will be conducted as a face to face, verbal interview. I have chosen the CAPI format for this study because this is the most effective mode of data collection for Spanish speakers living in linguistically isolated households in the U.S. (McGovern, 2004). In addition, the majority of Census Bureau surveys do not exist in multiple languages in paper form, and therefore CAPI interviewing is the most relevant mode of interview completion for many non-English speakers.

These interviews will include a debriefing section in which respondents will be asked their opinions on some of the cognitive interview techniques. The end product of this phase of the research will be a summary of the difficulties I encountered and a list of possible alternative techniques based upon both my findings and upon anthropological, sociolinguistic and survey methods literature.

3.3 Research Phase 2

The second phase of the project will involve the convening of a focus group of 4-5 bilingual (Spanish/English) expert cognitive interviewers. I will present the group with my findings from the first phase of the project, along with the list of possible alternative techniques that I have identified. I will seek to find out whether the experts have encountered the same issues while conducting interviews in Spanish or English, and ask them to describe adaptations that they have used successfully in the past. Finally, I will ask for their feedback on the alternative methods and solicit additional recommendations. Using the findings from the preliminary interviews and the focus group, I will construct a new cognitive interview protocol for the same series of American Community Survey questions.

3.4 Research Phase 3

Finally, using the new protocol, I will conduct 20 Spanish-language cognitive interviews with new respondents of similar demographic characteristics to those in Phase 1 in order to explore the use of the alternative techniques that we have identified. See Table 2 for a summary of the respondents' characteristics for Phase 3 of the study. I will then examine both the reactions of respondents to the cognitive interview probes and the type of information that this new type of interview elicits from them.

Table 2 Summary of respondent characteristics for Phase 3 of the research

	Spanish speakers
Low educational level (-12 years)	10
High educational level (12+ years)	10
Total respondents	20

This exploratory research will offer insight into methods for the development of appropriate techniques for conducting cognitive interviews for use in the development of multilingual and multicultural survey instruments.

4 Research Currently Underway at the U.S. Census Bureau

4.1 Changes to the research plan

Since I presented the above research plan at the Third International Workshop on Comparative Survey Design and Implementation in March of 2005, the project has evolved in a number of ways and the research is currently underway. First of all, I have reduced the scope of the study to include only one research phase. The research is now comprised of the conduct of a total of 40 exploratory cognitive interviews; 20 with monolingual Spanish-speaking respondents and 20 with monolingual English-speaking respondents. I have included English-speaking respondents in the study in order to compare the efficacy of common cognitive interview techniques across language/cultural groups and across educational levels. See Table 3 for a summary of my respondents' characteristics.

Similar to the original proposal, the cognitive interviews are based on a segment of the CAPI version of the U.S. Census Bureau's American Community Survey. The interview protocols now include an extensive debriefing section in which I collect information about respondents' backgrounds and experience with surveys. This enables me to create a case study of each respondent, in order to examine differences in their experience and comfort level with the survey process.

4.2 Nationality/ethnicity of respondents

In order to reduce linguistic variability due to national origin and ethnicity, I have restricted my sample to Mexican immigrants and native born (U.S.) non-Hispanic, white English-speakers. All interviews are being conducted in Texas, West Virginia and the Washington, D.C. metro area.

Table 3 Summary of respondent characteristics for current research

	English speakers	Spanish speakers
Low educational level	10	10
High educational level	10	10
Total respondents	20	20

4.3 Research questions

My current research examines cognitive issues, status issues, and sociolinguistic issues, and includes a number of research questions. Examples of my research questions are listed below.

Cognitive issues

1. What kind of introduction works best to explain the purpose of the cognitive interview to Spanish and English-speaking respondents of different educational levels?

Status issues

2. How do perceived status differences between the interviewer and the respondent affect rapport between the two parties?
3. What are the most effective ways to minimize perceived differences in status between interviewer and respondent?

Sociolinguistic issues

4. Does the cognitive interview interaction go more smoothly when the interviewer engages in small talk with the respondent for a time before beginning the interview? Is this true for some respondents and not others?
5. What types of respondents have more difficulty with common probe wording?
6. Are there alternative probes that work best with specific types of respondents?

Implications for future non-English language cognitive interview projects

7. Do these findings suggest any information about the type of training that would be useful for new cognitive interviewers who will conduct interviews in non-English languages?

4.4 Evaluation of results

All interviews are being tape recorded and transcribed, and I am doing systematic text analysis to evaluate the results. I am coding interviews for things such as whether a given probe or explanation was understood by the respondent the first time he/she heard it, whether a probe caused respondent discomfort, whether probes needed to be rephrased for respondent comprehension, whether the respondent was able to answer the probe in a manner consistent with interviewer expectations, and whether the probe elicited useful information for evaluation of the survey question.

4.5 Results

On the whole, the results of this research should add to our “tool kit” of cognitive interview techniques and should aid in the development of cognitive interview protocols in Spanish and other non-English languages. I will be presenting preliminary results of this research at the Federal Committee on Statistical Methodology Research Conference, November 14-16, 2005 in Washington, D.C.

5 Future Applications of the Present Research

This exploratory research is not expected to provide conclusive results, but it will offer insight into methods for the development of appropriate techniques for conducting cognitive interviews for use with multilingual survey instruments. Future applications of this research include the expansion of these methods to the conduct of cognitive interviews with survey instruments in additional languages and to the use of other pretesting methods such as focus groups, expert review and behavior coding for the development and testing of multilingual instruments.

References

- Bickart, B., & Felcher, E. M. (1996). Expanding and Enhancing the Use of Verbal Protocols in Survey Research. In N. Schwartz & S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research* (pp. 115-142). San Francisco: Jossey Bass.
- Blumberg, R. L., & Goerman, P. L. (2000). *Family Complexity among Latino Immigrants in Virginia: An Ethnographic Study of their Households Aimed at Improving Census Categories*. Report submitted to the United States Census Bureau.
- Briggs, C. (2001). *Assessment Report: Language Program*. Unpublished Census Bureau Report (November 20, 2001).
- Carrasco, L. (2003). The American Community Survey (ACS) en Español: Using Cognitive Interviews to Test the Functional Equivalency of Questionnaire Translations. *Study Series: Survey Methodology* #2003-17. Methodology and Standards Directorate, U.S. Census Bureau.
- Coronado, I., & Earle, D. (2002). *Effectiveness of the American Community Survey of the U.S. Census in a Borderlands Colonia Setting*. Draft report submitted to the U.S. Census Bureau, March 19, 2002.
- Gerber, E. (2004). *U.S. Census Bureau Official Cognitive Interview Training Course*. (Unpublished materials).

- Gerber, E. & Pan, Y. (2004). *Developing Cognitive Pretesting for Survey Translations*. Paper presented at the Second International Workshop on Comparative Survey Design and Implementation (CSDI). April 1-3, 2004. Paris, France.
- Harkness, J. A. (2004). *Problems in Establishing Conceptually Equivalent Health Definitions Across Multiple Cultural Groups*. Paper presented at the Eighth Conference on Health Survey Research Methods. Cohen, S. B., & Lepkowski, J. M. (Eds.) Hyattsville, MD: National Center for Health Statistics, 2004.
- Harkness, J. A., Van de Vijver, F. J.R., & Mohler, P. Ph. (Eds.) (2003). *Cross-Cultural Survey Methods*. John Wiley & Sons: Hoboken, New Jersey.
- Hunter, J., & Landreth, A. (2005). *Behavior Coding Analysis Report: Evaluating Bilingual Versions of the Non-Response Follow-Up (NRFU) for the 2004 Census Test*. U.S. Census Bureau report, submitted February, 18, 2005.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (1984). *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Report of the Advanced Seminar on Cognitive Aspects of Survey Methodology. Washington, D.C.: National Academy Press.
- Kissam, E., Herrera, E., & Nakamoto, J. M. (1993). *Hispanic Response to Census Enumeration Forms and Procedures*. Contract Report submitted to the U.S. Census Bureau, March, 1993.
- McGovern, P. D. (2004). A Quality Assessment of Data Collected in the American Community Survey (ACS) from Households with Low English Proficiency. Study Series: *Survey Methodology* #2004-01. Methodology and Standards Directorate, U.S. Census Bureau.
- McNally, J. W. (2000). *Overweight and Underexposed: The Measurement of Non-English Speaking Populations in the U.S. Representative Surveys*. Paper prepared for the National Institute of Child Health and Human Development Institute on Aging Conference on Inclusion of Language-Minority Populations in National Studies: Challenges, Opportunities and Best Practices. Bethesda, MD, July 27-28, 2000.
- Miller, K. (2003). Conducting Cognitive Interviews to Understand Question-Response Limitations among Poorer and Less Educated Respondents. *American Journal of Health Behavior*, 27(3), 264-272.
- Pan, Y. (2004). *Cognitive Interviews in Languages Other Than English: Methodological and Research Issues*. Paper presented at the American Association for Public Opinion Research conference, May 13-16, 2004, Phoenix, AZ. 2004 Proceedings of the American Statistical Association (to appear).
- Pan, Y. (2003). *The Role of Sociolinguistics in the Development and Conduct of Federal Surveys*. Proceedings of the Federal Committee on Statistical Methodology Research conference. November 17-19, 2003, Arlington, VA.

- Potaka, L., & Cochrane, S. (2004). Developing Bilingual Questionnaires: Experiences from New Zealand in the Development of the 2001 Maori Language Survey. *Journal of Official Statistics*, 20(2), 289-300.
- Potaka, L., & Cochrane, S. (2002). *Developing Bilingual Questionnaires: Experiences from New Zealand in the Development of the 2001 Maori Language Survey*. Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods (QDET), Charleston, SC, November 14-17, 2002.
- Presser, S., Rothgeb, J., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (Eds.). (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley & Sons, Inc.
- Rogler, L. H. (1999). Methodological Sources of Cultural Insensitivity in Mental Health Research. *American Psychologist*, 54(6), 424-433.
- Shin, H. B., & Bruno, R. (October 2003). *Language Use and English-Speaking Ability: 2000*. Census 2000 Brief. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf> on 2/15/05.
- Sirken, M. G., Herrmann, D. J., Schechter, S., Schwartz, N., Tanur, J. M., & Tourangeau, R. (Eds.) (1999). *Cognition and Survey Research*. New York: John Wiley & Sons, Inc.
- United States Census Bureau (2004). Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials. U.S. Census Bureau, U.S. Department of Commerce, Washington, D.C.
- United States Census Bureau (2003). *Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses*. Methodology and Standards Directorate, U.S. Census Bureau, U.S. Department of Commerce, Washington D.C., July 25, 2003.
- Wellens, T. (1994). The Cognitive Evaluation of the Nativity Questions for the Current Population Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 204-1209.
- Willis, G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.
- Willis, G. (2004). *Overview of Methods for Developing Equivalent Measures Across Multiple Cultural Groups*. Paper presented at the Eighth Conference on Health Survey Research Methods. Cohen, S. B., Lepkowski, J. M. (Eds.) Hyattsville, MD: National Center for Health Statistics, 2004.

CROSS-CULTURAL COMMUNICATION AND THE TELEPHONE SURVEY INTERVIEW

BRIAN KLEINER & YULING PAN

Survey researchers have long recognized the interactive, collaborative nature of orally administered survey interviews and how, as a type of talk-in-interaction, they differ from ordinary conversation (e.g., Suchman & Jordan, 1990; Schaeffer, 1991; Houtkoop-Steenstra, 2000; Maynard & Schaeffer, 2002). By focusing on the interactive and collaborative aspects of orally administered survey interviews, this body of research has illustrated how such a view not only helps to understand the survey process, but also opens the way for improvement in questionnaire design, testing, and interviewing techniques.

While the interactive nature of face-to-face and telephone survey interviews is widely recognized by survey researchers, there appears to be less appreciation of the fact that communicative norms for interaction in orally administered surveys vary widely from one culture and language group to the next, which may have significant consequences for survey non-response and data quality. We would like to make the case that telephone surveys that cross cultural and linguistic borders must take into consideration not only lower level semantic and pragmatic equivalence, but also must be sensitive to cross-cultural variation in the norms of language use. What is needed is a framework for understanding and documenting differences in communicative norms that may arise at different points in telephone surveys conducted across cultural and linguistic borders. This paper will identify the parameters of variation within such a framework that should be considered by survey researchers and translators when crafting telephone surveys.

1 Note on the Translation of Surveys

Current best practices and guidelines for questionnaire translation ultimately aim to promote equivalence of meaning and reference between questions in a source survey and those in translated counterparts (Harkness, 2003). Unfortunately, the guidelines followed in the field focus mostly on lower level semantic and pragmatic equivalence, and very little on higher level discourse features. Harkness, Van de Vijver & Johnson (2003) dis-

cuss two main approaches to translation from a source survey to other languages. The “adopt” approach involves strict or close translation of survey items in order to maintain the measurement properties of source items. The “adapt” approach involves increasing cultural appropriateness by modifying questions. While the need for closeness of translation is respected by Harkness (2003), she also points out that there may be linguistic and pragmatic differences across languages that require divergence from close translation. She also notes that survey translations that do not pay heed to these differences may be compromised:

“It is misguided to believe that we can develop questions with a unique and literal reading based on the lexical meaning of words alone. Thus, survey translations that focus on conveying lexical meaning and exclude consideration of pragmatic meaning have an obvious Achilles heel....To date, there has been little consideration of pragmatics research in literature dealing with questionnaire translation” (p. 48).

Harkness illustrates this point with examples limited to sentence-level pragmatic differences across languages as they relate to survey questions, such as grammatical gender and the use of answer scales. The remainder of the current paper will show that the Achilles heel extends to higher level differences across languages, which also must be heeded in translating surveys, especially those that involve the interaction of interviewers and respondents.

2 The Telephone Survey as a Speech Event

The telephone survey is a subtype of the interview speech event, which is governed by specific norms of language use. Unlike ordinary conversation, the telephone survey is highly structured with its own special characteristics. First, its participants are usually strangers, with one playing the role of interviewer and the other respondent. And while the mutual goal of interviewer and respondent is to arrive at accountable answers, each is subject to different constraints. While the interviewer initiates new topics and asks questions, the respondent is limited to answering and may ask questions only for clarification or to inquire about the overall management of the interview. The telephone survey has its own organization, with an opening, a body, and a closing, each with its own unique procedures and norms. Most importantly, the survey interviewer is subject to the constraints of standardization, meaning that questions must be read as worded (to obtain an answer that can be seen to be produced by the respondent and not the interviewer), probes must be done nondirectively, and a cooperative relationship must be maintained for the duration of the interview. Standardization aims for consistency of interview behavior, so that answers are comparable.

At the local level of interaction, with respect to the survey interview, the proper recording of accountable answers for social scientific purposes is an achievement that depends on an organization of interviewers' and recipients' concerted practical actions. These include specific kinds of speech acts, such as questions, answers, clarifications, probes, etc., and have been referred to in the literature as the "interactional substrate" (Maynard & Schaeffer, 2002). Research has shown that examination of interactional detail permits the identification of conversational practices – elements of the interactional substrate – that persist in the relatively constrained environment of the standardized telephone survey.

The telephone survey has organizational features of both interviews and telephone conversations. The telephone survey is sharply bounded by opening and closing sections, each conducted with respect to culture-specific norms. Openings of telephone surveys are far more elaborate than openings of ordinary telephone conversations. Opening sections of telephone surveys generally involve a variety of interactional tasks, including greetings, identifying the caller, establishing the relationship of the caller to the respondent, explaining the purpose of the call, determining the appropriate respondent, and addressing any concerns of the respondent. The opening of the telephone survey sets the stage for the interview section, defining roles and expectations for the interaction that follows. The way the opening unfolds between the interviewer and respondent may have significant implications for both survey non-response and data quality (Couper & Groves, 2002). Closings of survey interviews are usually more abbreviated than closings of ordinary telephone conversations, and most likely have no impact on survey non-response and data quality.

The opening section of the telephone survey is followed by the interview proper, which consists of sequences of questions and answers, interspersed with new topic introductions and side sequences in which interviewers and respondents diverge from the script to collaborate in arriving at accountable answers. Such side sequences may involve requests for clarification or elaboration, or repairs of misunderstandings (Moore & Maynard, 2002; Suchman & Jordan, 1990).

3 Cross-Cultural Communication and the Telephone Survey Interview

While there is little existing research on variation in communicative norms in telephone survey interviews that cross cultural or linguistic borders, much has been done examining how telephone conversations and communicative norms more generally vary across cultures. The examples that follow draw on this body of research and illustrate how telephone surveys, as a specific type of talk-in-interaction, are subject to a variety of cross-cultural differences. We contend that these differences are not trivial, and that they may affect both survey non-response and data quality.

3.1 Openings

Research has revealed that the norms governing openings of telephone conversations vary considerably from one culture to the next. For example, the practice of immediate identification by name is standard telephone practice in the United States, but is uncommon in Chinese language environments. Pan, Scollon & Scollon (2002) found that Chinese sales representatives in an American training program resisted the suggested practice of self-identifying when they made a phone call or as soon as they answered the phone. Dutch interactants, in contrast tend to self-identify by name when they answer the telephone, as Houtkoop-Steenstra (1991) found in an examination of opening sequences of 87 Dutch telephone calls. This behavior was interpreted as a response to a summons (i.e., the telephone ringing), which was usually followed by a reciprocal self-identification by the caller in the ensuing turn. According to Houtkoop-Steenstra, the Dutch response may reflect a cultural orientation toward the local accomplishment of social status, something less likely to be found in American telephone calls. Lindstrom (1994) compared the sequential organization of Swedish telephone conversation openings with those in the United States and the Netherlands. She concludes that

“in the United States there is a preference for other-recognition over explicit self-identification. In the Netherlands, by contrast, this preference is reversed....(t)he Swedish preference lies between the American and Dutch extremes” (p. 231).

Pan, Scollon & Scollon (2002) looked at reactions to three business telephone calls conducted by a Hong Kong professional. While Hong Kong focus group participants viewed the relatively direct openings of the professional as fairly natural and matter of course, Beijing focus group participants reacted quite negatively to the three calls, viewing them as too “business-oriented” and lacking in “personal feeling,” since the caller moved too quickly into the main point of the call. They felt that the pace was too fast, and the conversation was uncomfortably brief and abrupt. They commented that the fast speed and direct topic introduction gave the impression of a cold “business-is-business” call. Although they themselves sometimes receive these no-nonsense calls, they did not appreciate them. To these people, who are mostly in the same business, the caller sounded busy and indifferent. They said that he “didn’t care whether the other party was listening or not,” “didn’t try to get the attention of the other side,” and “didn’t have any emotion.” They concluded that this was the Hong Kong style, which represented a busy, modern society that lacked interpersonal interaction and close relationships.

Similarly, in a study comparing Greeks and Germans, Pavlidou (2002), found that Greeks prefer an exchange of phatic utterances before arriving at the reason for a telephone call, whereas Germans tend to opt for a more direct path to the main point of the call. These examples show that telephone survey researchers and translators conducting surveys

across cultural or linguistic borders should be sensitive to such differences. While quickly coming to the point of a call may be appreciated in one culture, it may cause offense in others, potentially leading to survey non-response.

3.2 Questions, answers, and topic shifts

While questions and answers are most likely part of the communicative repertoire of all cultures and language groups, the way they are carried out may vary from one culture to the next, with implications for telephone survey interviews that cross cultural or linguistic borders. For example, questions that call out a dispreferred response may give rise to different behaviors across cultures (within conversation analysis, dispreferred responses are those that are presumed by the answerer to be contrary to the hopes or expectations of the questioner). In many Asian cultures, under certain circumstances, people tend to avoid “no” answers in response to yes/no questions. This is especially the case where there is an asymmetrical relationship between interlocutors in more formal types of interaction. Within telephone survey interviews, avoidance of dispreferred “no” responses may be highly problematic, requiring at the least considerable work on the part of interviewers to ferret out a definite “yes” or “no” answer.

In many or perhaps most cultures, answers to questions are obligatory. Questions that cannot be answered are generally accompanied by an explicit explanation (e.g., “I can’t answer that now”). Conversation analysts have noted that given the sequential organization of talk and the nature of adjacency pairs, anything immediately following a question will be interpreted as an answer. However, in some cultures, answers to questions are not obligatory. For example, Philips (1990) found that immediate answers to questions are not normative for the Indians of the Warm Springs Reservation in central Oregon. This has obvious relevance to telephone survey interviews, which rely on a shared norm that questions must be followed by answers.

Scollon & Scollon (1995) state that there is a western preference for a deductive pattern for topic introductions, whereas eastern cultures prefer an inductive pattern for introducing topics. They describe a typical call-answer-topic pattern for western cultures, in which the first topic is generally introduced immediately by the caller after establishing contact with the answerer. Asian cultures, in contrast, tend to exhibit a call-answer-facework-topic pattern, where the new topic is delayed until there has been an adequate display of facework on the part of the interlocutors. According to Scollon & Scollon, the purpose of the facework in this sequence is to allow both parties to get a sense of each other’s mood or position. Such differences in the norms for topic introduction may have repercussions for cross-cultural telephone surveys. For example, an Asian respondent might feel that an

American interviewer was too abrupt or even rude by skipping the requisite amount of facework, whereas an American might view the lead-up to the main topic as irrelevant or even annoying.

3.3 Global differences in communicative norms

The previous sections have addressed cross-cultural differences in communicative norms that may affect specific parts of telephone survey interviews, namely openings, question/answer sequences, and topic shifts. Other kinds of cross-cultural variation that may also have an impact on surveys are not limited to specific sections of the telephone survey. These include level of exposure to the survey interview as a type of speech event and establishment and maintenance of rapport with respondents.

Exposure to surveys. It should not be assumed that all people know how to participate appropriately in a survey interview. In fact, familiarity with surveys as a type of interview varies from one culture and subgroup to the next. For example, Pan & Scollon (2004) found that many Chinese immigrants in the United States lack any experience of participating in surveys, either in their home country or in the United States. In China, census data and other demographic statistics are collected through administrative channels, and individual citizens are not aware of the need or the procedure of collecting data through surveys. Of the 40 respondents interviewed, none had any experience with telephone surveys.

Hughes (2004) investigated how United States immigrants from three countries perceive surveys. She found that Chinese, Ethiopian, and El Salvadoran respondents (eight from each group) had no experience with surveys in their home countries, but that most had experience filling out school forms. In Ethiopia and in El Salvador, there is no official census. In China, the official census is conducted through work units (companies, organizations) and local community committees. Most of the respondents had heard about the U.S. 2000 Census through a massive advertising campaign, but few had participated in it. Many respondents knew about the Census telephone survey, but did not participate, because they felt uncomfortable giving out their information to a stranger over the phone. Clearly, variation in exposure to the survey interview speech event could have implications for the comparability of data collected across cultures.

Establishing and maintaining rapport. An important aspect of telephone survey interviews is the establishment of rapport and relationship between the interviewer and respondent. Problems arising from the failure of interviewers to behave in culturally appropriate ways can lead to awkwardness at the least, and survey non-response at the worst. Telephone surveys that are translated too closely from a source survey run the risk of

forcing interviewers to violate various cultural norms involving face, politeness, and the linguistic encoding of status and social distance.

Pan, Scollon & Scollon (2002) reported on a case study of an American telecommunications training program. The training manager was trying to encourage his Chinese-speaking sales representatives to use phrases such as “please” and “thank you” in making their sales pitch. The sales representatives were given a Chinese script translated directly from the English script. According to the script, the sales representatives were supposed to use “please” and “thank you” at every possible point in their conversation with Chinese-speaking customers, and their performance would be marked down if they failed to do so. During the three days of performance evaluation, 10 Chinese-speaking sales representatives were monitored. Among them, nine received zero points in the category of politeness level, because they failed to use “please” and “thank you” in their conversations with Chinese customers. The tenth person received one point out of three for this category. As a result, they all failed in their performance evaluation, because their politeness level was not up to the company’s standards. The irony is that the calls monitored were very polite by Chinese standards, and most of the sales representatives succeeded in signing up their customers for the company’s service. The sales representatives were very frustrated with the training program. One woman even yelled, “Mark me down if you will. I cannot do it.”

In many languages, including Chinese, it is not natural to use expressions like “please” and “thank you” in daily conversations. These terms are used mostly on very formal occasions; in other situations, they imply a large social distance between two speakers. When sales representatives use these polite terms excessively in their conversation with customers, they may sound suspicious to Chinese people, who may feel they are being trapped into buying some service. So despite the rigid training, the Chinese-speaking sales representatives still refused to use “please” and “thank you,” instead preferring other forms of polite expressions. In Chinese, for daily conversations, politeness is indicated by other linguistic features, such as repetition of a verb (e.g., “take, take,” “see, see,” “read, read”); repetition of adjective phrases (e.g., “good, good,” “fine, fine”); a tag question (e.g., “Is that OK?”); tone of voice, intonation, rate of speech, or the insertion of appropriate pauses; and other discursive features.

Another global consideration is the encoding of relationship between the interviewer and respondent in telephone surveys. In some languages, for example, Korean and Japanese, aspects of relationship such as relative power and social distance are encoded at lower linguistic levels in the grammar. Surveys that are translated into such languages must allow for multiple linguistic choices on the part of interviewers, depending on the identity

of the respondent. However, since most interviewers and respondents are strangers, and since the visual cues for identity are absent during the telephone survey interaction, some initial interactional work at the beginning of the interview may be needed to establish a working relationship for linguistic purposes. Translated surveys that do not allow for this initial work run the risk of offending respondents when interviewers choose inappropriate forms based on poorly informed assumptions about the identity of respondents (e.g., their age and status).

Forms and manners of address also vary considerably from one culture to the next, and direct translation of address terms from a source survey may lead to adverse consequences. While some cultures assume an ethos of closeness and solidarity between strangers, others assume greater social distance that must be reflected in terms of address. For instance, in Polish culture, it is entirely inappropriate to address someone other than a family member or friend by first name, whereas this is more acceptable in American culture. Addressing a Polish respondent by first name in a telephone survey would almost certainly result in non-cooperation. In Chinese culture, it is not uncommon to address a higher status stranger as “Uncle” or “Auntie” to display deference. Obviously, translation of a source survey in Chinese to some other language might need to modify or omit these special terms of address.

4 Discussion

No less than telephone conversations or face-to-face interviews, telephone survey interviews are a type of talk-in-interaction governed by culture- and language-specific norms. As such, telephone survey interviews that cross cultural or linguistic borders must be designed to take into consideration varying norms of language use. For surveys that are translated, this means that translations must be modified to fit the norms of the target language. Standard translation techniques might overlook these kinds of variation, preferring to translate more directly from the source language, possibly leading to various kinds of interactional troubles.

The examples provided earlier to illustrate this point indicate that there are both local parameters of variation and global considerations that survey researchers and translators must address. Telephone surveys that cross cultural or linguistic borders should take into consideration variation in norms of language use with respect to openings of calls, question and answer sequences, and topic shifts. They should also be designed and translated with sensitivity to global differences in norms of language use, including cultural experience with surveys and cultural differences in establishing, maintaining, and encoding rapport and relationships.

The notion that survey translations must sometimes diverge from a source survey in order to respect varying norms of language use appears to conflict with the aim of asking questions in an equivalent way, thus ensuring comparability of data. There is a paradox, therefore, with respect to conducting surveys across cultural and linguistic borders – close translation of survey items may in some cases compromise comparability of data.

Some of the problems resulting from a lack of sensitivity to cross-cultural differences in communicative norms may simply lead to awkwardness in interviews. An extreme result might be survey non-response or incomplete interviews. Others may lead to compromised data quality and affect measurement error. Future work should be empirical in nature and should examine which types of variation are more likely to increase measurement error and which are more likely to have more benign consequences.

References

- Couper, M. P., & Groves, R. M. (2002). Introductory Interactions in Telephone Surveys and Nonresponse. In D. W. Maynard et al (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 161-77). New York, NY: John Wiley and Sons, Inc.
- Harkness, J. A., Van de Vijver, F. J.R., & Johnson, T. P. (2003). Questionnaire Design in Comparative Research. In J. A. Harkness, F. J.R. Van de Vijver & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 19-34). Hoboken, NJ: John Wiley and Sons, Inc.
- Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. J.R. Van de Vijver & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35-56). Hoboken, NJ: John Wiley and Sons, Inc.
- Houtkoop-Steenstra, H. (2000). *Interaction and the Standardized Interview: The Living Questionnaire*. Cambridge: Cambridge University Press.
- Houtkoop-Steenstra, H. (1991). Opening Sequences in Dutch Telephone Conversations. In D. Boden, & D. H. Zimmerman (Eds.), *Talk and Social Structure. Studies in Ethnomethodology and Conversation Analysis* (pp. 232-50). Cambridge: Polity Press.
- Hughes, K. (2004). *Immigrants' Perception of the 'Survey': A Cross-cultural Comparison*. Research working paper, Statistical Research Division, US Census Bureau.
- Lindstrom, A. (1994). Identification and Recognition in Swedish Telephone Conversation Openings. *Language in Society* 23, 231-52.
- Maynard, D. W., & Schaeffer, N. C. (2002). Standardization and its Discontents. In D. W. Maynard et al. (Eds.) *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 3-45). New York, NY: John Wiley and Sons, Inc.

- Moore, R. J., & Maynard, D. W. (2002). Achieving Understanding in the Standardized Survey Interview: Repair Sequences. In D. W. Maynard et al. (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 281-311). New York, NY: John Wiley and Sons, Inc.
- Pan, Y., Scollon, S. W., & Scollon, R. (2002). *Professional Communication in International Settings*. Malden, MA: Blackwell Publishers.
- Pan, Y., & Scollon, S. W. (2004). *Opening the Circumference of Census Taking: A Nexus Analysis*. Paper presented at the 2004 Annual Conference of the American Association for Applied Linguistics. Portland, Oregon, May 1-4, 2004.
- Pavlidou, T. (2002). Moving towards Closing: Greek Telephone Calls Between Familiars. In K. K. Luke & T. Pavlidou (Eds.), *Telephone Calls: Unity and Diversity in Conversational Structure across Languages and Cultures* (pp. 201-29). Amsterdam: John Benjamins.
- Philips, S. U. (1990). Some Sources of Cultural Variability in the Regulation of Talk. In D. Carbaugh (Ed.), *Cultural Communication and Intercultural Contact* (pp. 329-344). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Schaeffer, N. C. (1991). Conversation with a Purpose - or Conversation? Interaction in the Standardized Interview. In P. P. Biemer et al. (Eds.), *Measurement Errors in Surveys* (pp. 367-392). New York, NY: John Wiley and Sons, Inc.
- Scollon, R., & Scollon, S. W. (1995). *Intercultural Communication*. Cambridge, Mass: Blackwell.
- Suchman, L., & Jordan, B. (1990). Interactional Troubles in Face-to-Face Survey Interviews. *Journal of the American Statistical Association* 85, 232-241.

THE USE OF INTERPRETERS IN THE CONDUCT OF HOUSEHOLD SURVEYS

Development of U.S. Census Bureau Interpretation Guidelines

YULING PAN

1 Introduction

According to the recent figures in census data, there is clearly an increasing need for non-English language data collection instruments and other survey materials in the United States. The Census 2000 Supplementary Survey revealed that, in the United States nationwide, there are approximately 45 million people aged 5 years and older who speak a language other than English at home. This represents about 18 percent of persons in this age group. Of these 45 million people, over 10.5 million speak English either “not well” or “not at all” (Li, et al., 2001). In addition, the U.S. Census Bureau reported that in 2000, 4.4 million households encompassing 11.9 million people were linguistically isolated. According to the Census Bureau’s definition, a linguistically isolated household is “...one in which no person aged 14 or over speaks English at least ‘very well.’” A linguistically isolated person is defined as “...any person living in a linguistically isolated household” (Shin & Bruno, 2003). The numbers of linguistically isolated households and people in the 2000 Census were significantly higher than those in the 1990 Census, when 2.9 million households and 7.7 million people were reported to be linguistically isolated (Shin & Bruno, 2003).

In order to meet the increasing demand for obtaining high quality data from the increasingly multi-lingual and multi-ethnic universe of respondents, the U.S. Census Bureau translated the 2000 decennial census questionnaire into five non-English languages (Spanish, Chinese, Korean, Vietnamese, and Tagalog) and provided language guides in 49 languages. In April 2004, the U.S. Census Bureau issued its translation guideline “*Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials*” to provide guidance on translation approaches and procedures to ensure the high quality translation of survey instruments and supporting documents.

However, obtaining high-quality data from households where English is not the home language requires more than just having a correctly translated data collection instrument. The Census Bureau decennial census and ongoing demographic surveys entail personal interviews. Personal interviews are conducted largely by monolingual, English-speaking field interviewers, and sometimes by bilingual field interviewers in some languages. There are two issues that call for our attention when interviewing non-English-speaking households. One is the use of interpreters in such an interview. When a monolingual English-speaking field interviewer encounters a household in which the adult members speak little or no English, he or she must rely on someone (an interpreter) who speaks the target language of the respondent for assistance in conducting the interview. There is a lack of standards and procedures for identifying interpreters to interpret an interview. The other issue has to do with the use of bilingual field interviewers who conduct oral translation on the fly. Bilingual field interviewers translate the script of a questionnaire while they interview respondents. We need to consider what level of bilingual competence and interpretation skills are required for bilingual field interviewers to effectively conduct the interpretation of a survey interview.

While there is growing research interest in survey translation methodology, its impact on data quality and the recommended approach for survey translation (e.g., Harkness, 2003; Harkness, et al., 2003; de la Puente & Pan, 2003), not much attention has been paid to the ways in which interpreters are currently being used in the conduct of household surveys. There is little research done on the use of interpreters and the interpreter's role in survey interviews. Standards and procedures for the use of interpreters in field interviews are lacking, and there is an urgent need for the development of such guidance.

Therefore we are proposing a research plan to develop guidance on how to best train and use interpreters in in-person interviews at the U.S. Census Bureau. This paper outlines the research plan and the proposed research activities to develop a Census Bureau interpretation guideline. It includes 1) goals of the research plan and research activities, 2) preliminary research and findings, 3) issues to be addressed in the proposed Census Bureau interpretation guideline, and 4) further research.

2 Research Plan

Our proposed research plan has the following goals:

1. To identify issues and challenges involved in conducting interviews with non-English-speaking respondents (i.e., interviews conducted in English using interpreters and in languages other than English by bilingual field interviewers);
2. To suggest ways to address the observed challenges associated with the conduct of in-person interviews using interpreters;
3. To investigate the current practices and procedures of assessing language proficiency of bilingual field interviewers and interpreters;
4. To assess what training is needed for interpreters and field interviewers to effectively interview non-English-speaking households.

The research plan includes the conduct of three main research activities. First, we will conduct a web search to identify interpretation guidelines used by research organizations and survey firms in the United States and around the world. The purpose of this effort is two-fold: one is to investigate current best practices for the use of interpreters in field interviews by survey organizations; the other is to determine whether there are any policies/guidelines about using interpreters.

The second research effort is to observe U.S. Census Bureau field interviewers conducting interviews with non-English-speaking respondents. The goal of the field observation is to better understand current practices and issues in interviewing non-English-speaking respondents through the use of interpreters and/or bilingual field interviewers.

The third research effort is to observe Census Bureau field interviewer training sessions, to review field interviewer training materials, and to conduct focus groups and debriefing interviews with field interviewers and other Census Bureau field staff. Through this research effort, we will obtain field interviewers' views of how using interpreters works for them and what guidance is needed in the interpretation guideline.

Sociolinguistics will be the principal paradigm used to analyze the data from the research and to generate recommendations. By this we mean that we will look into the characteristics of the social setting of the survey interview, the role of the interpreter in this kind of social setting, and requirements of linguistic and cultural competence for interpreting in survey interviews. The findings from this research will ultimately be used to develop Census Bureau guidelines for the use of interpreters in the conduct of household surveys.

3 Preliminary Research

Of the aforementioned research efforts, we have started two research activities. One is to conduct a web search of interpretation guidelines used by research organizations and survey firms in the U.S. and around the world. The other is to observe Census Bureau field interviewers conducting interviews with non-English-speaking households. The following subsections report the preliminary findings from these two research activities.

3.1 Preliminary findings from a web search of interpretation guidelines

Between September and November of 2004, we conducted a web search for guidelines on using interpreters, adopting the same search method as that used in a previous web search for translation guidelines (de la Puente, et al. 2004). Google was the primary search engine. Search engines available within specific websites were also used.

The web search consisted of two phases. In the first phase, we began searching the web for any existing interpretation guidelines used by research and social science organizations within the United States. These organizations include educational institutions, court systems, and health care organizations. Results indicate that some research organizations have guidelines for the use of interpreters. The court system has codes of conduct for interpreters, but few standards of practice exclusively addressed court and legal interpreters. Health care service delivery agencies have more established standards of practice for interpreting and provide more and better guidance than agencies and organizations whose work is not focused on health care delivery.

In the second phase of our web search, we focused on exploring the websites of only survey research organizations. The results of the second phase of the web search indicate that unlike research-oriented organizations, most survey research organizations do not have interpretation guidelines. However, some sites do indicate that there are guidelines, but they are not posted on the web. Other sites indicate a need for their organization to have guidelines, but it is unclear whether or not they are working toward creating guidelines. In these cases, we made telephone calls or sent email messages to the organizations to find out how they deal with various languages they encounter. We asked them questions like: What languages do you encounter? Do you hire interpreters or do you rely on bilingual or multilingual field interviewers? Are there specific interpreter guidelines or standards for international research? What are the problems involved with international research?

The response of these organizations was extremely positive. Many companies admitted that they have no idea whether or not anything has or will be done about language barriers

and interpreter guidelines for their organizations. Others admitted that they had never considered the issue and they commended the Census Bureau for tackling it.

3.2 Observation of field interviews

During the months of August and September of 2004, we observed monolingual and bilingual field interviewers conducting interviews in the field and from a Census Bureau telephone center. We selected areas that have a high concentration of the non-English-speaking population in the United States for our observations.

From the field observation, we identified three main issues with the use of interpreters in field interviews. These issues have to do with identifying interpreters, interpreters' qualifications, and interpreters' performance.

The first challenge is to determine whether there should be formal procedures/standards in place for identifying interpreters. The current practice is that some Census Bureau regional offices provide interpreters (the Census Bureau has 12 regional offices). But in most cases, field interviewers find their own interpreters (from their own resources, connections with the neighborhood, or on the spot). Field interviewers have no way to assess an interpreter's qualifications or performance. When asked how they know if an interpreter is doing a good job, one field interviewer said that he judged it by the interaction between the interpreter and the respondent. If the interaction seemed to be smooth, he assumed that the interpreter was doing a good job. In our observation, we noticed that an interpreter could carry out a seemingly smooth conversation with the respondent without accurately translating what the field interviewer or the respondent were saying.

The second issue is associated with the qualifications of interpreters used in in-person interviews. The qualifications of interpreters vary a great deal. Some regional offices keep a pool of interpreters that they can use for major languages in the region. On many occasions, a neighbor or a young household member is asked to be the interpreter for an interview. In these cases, those who serve the role of an interpreter do not have formal training in interpretation nor do they have work experience in interpretation. On top of that, they are called into the job on short notice. They do not receive any training or briefing about the interpretation assignment to prepare them for the work.

The third challenge has to do with interpreters' performance. From our field observations, we identified the following problems with interpreters' job performance:

1. Interpreters are not always familiar with the survey on which they are working; often they have not been informed of confidentiality and privacy policies, and this can lead them to omit this information when speaking to respondents.

2. Interpreters often usurp the role of the interviewer: they construct their own questions and sometimes even provide answers for the respondent.
3. Interpreters sometimes take on an additional role in the interview: lecturing respondents, initiating sidetrack conversations with the respondent.
4. Most importantly, interpreters are not always giving an accurate translation of what the field interviewer says to respondents and vice versa. They sometimes add or omit information.

We concluded that these problems occurred due to the lack of standards in identifying interpreters and the lack of job training for them.

3.3 Use of bilingual field interviewers

For some languages, the Census Bureau has hired bilingual field interviewers. This looks like an ideal situation because the field interviewers are familiar with Census Bureau surveys and procedures. However, we notice that there are several issues worthy of our attention. First of all, there are various types of bilingual field interviewers. Bilingual field interviewers consist of heritage speakers (speakers who have learned and used the target language in a home environment only), second language learners, and recent immigrants. Their bilingual competence varies considerably. During our observations, we noticed that some bilingual field interviewers were not proficient in the target language, and some of them did not have a strong English proficiency either. Another issue is that bilingual field interviewers conduct interpretation on the spot. They may not have the necessary interpretation skills to effectively convey messages across the two languages. Research shows that being able to speak the two languages does not always equate to being able to conduct interpretation. Interpretation calls for a special skill set to successfully transfer meanings from one language to the other (Gentile, et. al, 1996).

This brings us to the next set of research questions. That is, what level of bilingual proficiency is required for bilingual field interviewers to successfully handle interviews in two languages? What level of interpretation skills are required to interpret a survey interview? What procedures do we need to have in place to verify or certify bilingual field interviewers' bilingual proficiency and interpretation skills? And what assessment tool do we need to develop in order to evaluate the bilingual competence and interpretation skills of bilingual field interviewers?

4 Issues to be Addressed in Census Bureau Interpretation Guidelines

Based on our initial web search and field observations, we believe that it is important to provide guidance and standards for the use of interpreters. These guidelines or standards will help to ensure that the data collected in an interview through the use of an interpreter are as reliable and accurate as the data collected in standardized interviews in English. Preliminary research indicates that Census Bureau interpretation guidelines need to address the following critical issues:

1. What are the most appropriate criteria for use in selecting interpreters?
2. What standard procedure should be in place for using interpreters in the field?
3. What type of training is needed for preparing interpreters?
4. What materials should be provided for interpreters?
5. When encountering a non-English-speaking household, what type of protocols or procedures should field interviewers follow in identifying an interpreter on the spot?
6. What should be the key elements/features of the protocol/procedure?
7. How can a field interviewer best implement these procedures?

5 Further Research

With these issues in mind, we plan to carry out further research to find solutions to the problems. We will continue our web search to learn more about best practices in the use of interpreters by survey organizations around the world. We will also extend our field observations of Census Bureau field interviewers in different regions in the United States.

In addition, we will conduct research to identify what type of training is needed for Census Bureau field interviewers concerning the use of interpreters and what guidance is needed in the interpretation guidelines. We intend to observe field interviewer training sessions, to review field interviewer training materials, and to conduct focus groups and debriefing interviews with field interviewers and other Census Bureau field staff.

One of the most challenging questions about the use of interpreters in in-person interviews is “How do we know if someone who claims to speak the target language is proficient in that language and has the skill to do the interpretation?” The U.S. Census Bureau currently does not have bilingual staff in all relevant languages to assess the language

proficiency or interpretation skills of potential interpreters. We foresee the need to develop a practical assessment tool of interpretation skill that can be used across languages.

We intend to pursue this effort by incorporating the U.S. federal government Interagency Language Roundtable's description of translation skill levels. The interagency language roundtable has recently issued a draft of skill level descriptions for translation (levels 1-5, with level 1 being the lowest and level 5 being the highest). This is a self-assessment tool for translators. We plan to use this concept to develop a series of questions that could be used for self-assessment of interpretation skills to be used by bilingual field interviewers and potential interpreters. This is the first step in developing a practical tool for language assessment.

6 Final Goal of the Research Project

The ultimate goal of this research is to develop a U.S. Census Bureau interpretation guideline, incorporating findings from the present research and best practices of survey organizations around the world.

We envision that the U.S. Census Bureau interpretation guideline will consist of the following components:

1. Guidelines on protocols, procedures, and standards for using interpreters in interviews;
2. Guidance on interpreter training (survey information, confidentiality, interpreters' role in a survey interview, codes of ethics and codes of conduct);
3. Development of a self-assessment tool of interpretation skill and/or bilingual proficiency of field interviewers and potential interpreters.

As mentioned in the beginning, this research project is still at an early stage. We believe that by undertaking this research effort, we will be able not only to develop a Census Bureau interpretation guideline, but also to identify issues relevant to the development of the interpretation guidelines for the International Workshop of Comparative Survey Design and Implementation.

References

- de la Puente, M., & Pan, Y. (2003). *An Overview of Proposed Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials*. Proceedings of the Federal Committee on Statistical Methodology Research Conference (Arlington, VA).
- de la Puente, M., Pan, Y., Gerber, E., & Cidade, M. (2004). *Preliminary Content Analysis of Translation Guidelines and Best Practices Available on the World Wide Web*. Paper Presented at the Second International Workshop on Comparative Survey Design and Implementation. Paris, France. April 1-3, 2004.
- Gentile, A., Ozolins, U., & Vasilakakos, M. (1996). *Liaison Interpreting: A Handbook*. Victoria, Australia: Melbourne University Press.
- Harkness, J. A. (2003). Questionnaire Translation. In J. A. Harkness, F. J.R. Van de Vijver & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 35-56). New Jersey: John Wiley & Sons, Inc.
- Harkness, J. A., Van de Vijver, F. J.R., & Johnson, T. P. (2003). Questionnaire Design in Comparative Research. In J. A. Harkness, F. J.R. Van de Vijver & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods*. New Jersey: John Wiley & Sons, Inc.
- Li, R. M., McCardle, P., Clark, R. L., Kinsella K., & Berch, D. (Eds.) (2001). *Diverse Voices Inclusion of Language-Minority Populations in National Studies: Challenges and Opportunities*. National Institute on Aging and National Institute of Child Health and Human Development. Bethesda, Maryland, USA.
- Shin, H. B., & Bruno, R. (2003). *Language Use and English-Speaking Ability: 2000*. Census 2000 Brief. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. Retrieved August 31, 2005, from <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>
- United States Census Bureau. (2004). *Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials*. Washington, DC: U.S. Census Bureau, U.S. Department of Commerce. Retrieved April 5, 2004, from [http://www.census.gov/cac/www/Paper\(3\)-LanguageWG_Spring2004.html](http://www.census.gov/cac/www/Paper(3)-LanguageWG_Spring2004.html)

CATEGORY AND COMPARISON ACROSS WHAT KIND OF FRONTIER?¹

JOHN MACINNES

1 Category and Comparison

Both variables and the values they take can be thought of as categories. Comparison across categories (e.g. in the sense of comparing the proportion of cases within a category of a variable) is the basic building brick of any social science methodology, including so-called ‘qualitative’ ones. What tends to change is simply the nature of the categories and how information about them is collected (or how variables are measured). A category can be thought of as boundary drawn around a set of items possessing a common characteristic; e.g. human beings who are males, people born in 1956, natives of Aberdeen, the currently married, parents with one child, possessors of higher degrees, non-voters in the last election, part-time workers, people employed in the public sector, think of themselves as equally Scottish, British and European, residents of Spain, inhabitants of Barcelona. These dozen categories probably uniquely identify me. Unfortunately, even crudely understanding or modeling my attitudes or behaviour would require dozens more.

2 The Specificity of ‘Comparative’ Research

If we understand social science research in this way, then the only distinctive feature of ‘comparative’ research is that one of the boundaries used to circumscribe a category is the frontier of a state territory. This is rather idiosyncratic. It is not intuitively obvious why comparisons, say, across different government regions or nations *within* a state territory (including nations in which the language normally used changes, or government regions with different legal, health or education systems etc.) are qualitatively different to comparisons *between* states (including states joined by language, policy contexts or historical evolution – e.g. Ireland and the UK – or states that are together part of some higher level economic and political unit such as the European Union).

1 This article was produced as part of research financed by the Plan Nacional I+D+ and Ministerio de Educación y Ciencia (Spain) (Ref. SEJ 2005-03764/GEOG).

3 The Social Construction and Interpretation of Categories and the Inevitability of the ‘Emic’

All categories are socially constructed, (as opposed to that which they represent which may or may not also be so constructed). As such they are by definition open to diverse interpretation. Some, however, are more socially constructed than others, insofar as some can be readily connected to physical or social processes that operate independently of the observer or categoriser, while others cannot. There is no better example of the first than that cited with typical humour by the late Ian Craib (1998):

while I might conceive of my liver in all kinds of exotic ways, these can never change how it actually functions. However, as medical knowledge develops ever more sophisticated categories that enable it to ‘model’ my liver’s operation, doctors might well come to ‘understand’ it well enough to look after it, repair or even replace it.

Age in years last birthday is a fairly straightforward set of categories, about which it is fairly easy to establish shared and agreed meaning between researcher, the subjects of the research and the audience(s) for the research results (at least for those with some conception of calendrical time).

Conversely, other categories can, by definition, lie only in the eye of the beholder, whether or not the latter is aware of this, because their existence cannot be held to ‘objective’ empirical test. Gender identity, class or employment characteristics, for example, are likely to generate discussion about what categories might or might not be appropriate or even exist at all, and how they ought best to be defined and understood.

Linguistics, following Pike (1954) has a useful terminology for these two forms of category: the former are *etic* and the latter are *emic*. I think it is often useful to further divide the ‘emic’ category to specifically identify those categories that suffer not only disagreement about their definition, but whose very existence in the first place stems from the logical impossibility of defining them. (*Why* such categories might exist lies beyond the scope of this paper. Essentially the argument is that such categories allow people to simultaneously assert and deny category membership by discursively camouflaged shifts in definition. I have made such a case for the category ‘gender’ in MacInnes (1998b). Try, for example, defining ‘masculinity’. The category depends on a connexion to the category ‘male’ that must also be denied for the category to have any sense in the first place. Goffman gives a good example when discussing ‘masculinity’:

There is only one complete unblushing male in America: a young, married, white, urban, northern, heterosexual Protestant father of college education, fully employed, of good complexion, weight, and height, and recent record in sports. (1963: 128)

Translating his argument into our terminology we could simply say that no American or any other male corresponds to the category 'masculine' as comprehensively defined.

A prime example of this second type of emic category is that of 'nation'. Academic observers are happy to admit (whether or not they think that 'nations' exist) that no comprehensive definition of the category is possible. Not only is there routine disagreement about what might constitute the boundaries of a particular or any nation, but also about what class of phenomena nations are, or indeed whether they 'exist' at all. The existence of the category thus serves less to classify objects, than to obscure the fact that they *cannot* be classified. As we shall see, this can lead both survey researchers and others into trouble if they then attempt to use these pseudo categories in the normal way. I shall argue below that the concept of a 'society' empirically bounded by a state frontier, is very close to, if not synonymous with, such a category of 'nation'.

4 Shared Understanding of Categories

A common problem in all social research is that different 'categories' of respondents may have systematically distinct understandings of the meaning of the 'categories' that we wish to use in researching them. Unless we can control this we will not be able to separate out the effects of variation in respondent interpretation of the categories used in the questions we ask, and variation in the substantive answers they give. However, although this is a problem routinely raised when conducting research using respondents from different states, frequently speaking different languages (so that a lot of effort goes into accurate translation across languages for example) it is present in research of any kind, because it arises from the nature of *any* respondent variation, not just variation in the language respondents speak. We might expect men and women, or the currently employed and others, or old and young to also to have distinct understandings of categories which vary systematically with these characteristics. The difficulty of controlling adequately for this is usually the rationale for ethnographic or more 'qualitative' approaches in the first place. Such approaches cannot 'solve' this problem, but they should make us more alert to issues of the variable interpretation of meaning in *any* context, not only those routinely understood as 'comparative' in the sense of comparisons between states.

5 States as Etic Categories

The specific methodological issue faced by comparative research is therefore the nature of the state frontier as a category and ways it might be interpreted. State frontiers are a resolutely 'etic' category in the sense that states jealously guard their territorial sovereignty over even quite useless bits of land. For example Spain and Morocco recently engaged in belligerent confrontation over a tiny Mediterranean island inhabited by a few goats.

Crossing state boundaries formally requires passports and visas. The writ of state specific law stops abruptly at the frontier and so on. Trying to convince an immigration official that the frontier they police lies only in their imagination, is open to diverse interpretations or is a matter for discussion would be a fruitless task.

However state frontiers are *emic* in the sense that as a result of wars, revolutions or other political change, they shift and develop over time. Indeed the category of 'sovereign state territory' itself is thoroughly modern, superseding the parcelisation of sovereignty by status and spatially overlapping jurisdictions characteristic of previous ages (see e.g. Anderson, B., 1992; Anderson, P., 1978; Billig, 1995; Bloch, 1939). States are hardly less mortal than human beings, witness the map of virtually any area of the world save North and South America over the last half century or so. This does raise, however, the question of what constitutes a 'state', or what our basic units of analysis in comparative research comprise. What fifteen years ago were the units of Czechoslovakia, the Federal Republic of Germany and the German Democratic Republic are now The Czech and Slovak Republics and the united Germany. If it makes sense to continue to treat former East and West Germany as separate units of analysis, would it not, with appropriate hindsight, have been equally logical to do the same with what were to become the Czech and Slovak republics? Would it make similar sense to do so with the warring remnants of the former Republic of Yugoslavia? Should we anticipate potential future developments and insist that Scotland, Euskadi, Catalonia, Quebec and Flanders are similarly always analysed separately? But why stop there? Closer scrutiny will reveal literally hundreds of separatist movements within modern 'established' states, some with real chance of political success, some the cherished vision of a handful of cranks.

If state frontiers shift across space (often dragging refugees in their wake) so does what they contain change over time. Ought we to treat as continuous and obvious categories, for example, the states of central and Eastern Europe before and after the revolutions of 1989? After Ceausescu's murder, Rumania became a different 'place' in a very real sense, without its geographical location changing a centimetre. How substantial does 'regime change' have to be to establish the existence of a different unit? Comparative analysis would be in trouble if we insisted that we treated states as distinct units each time their government changed. It might be equally misleading, however, to insist on territorial continuity in the face of dramatic political or social shifts. On election night in the UK in May 1997, more than one observer argued that the results were the poorest for the Conservative Party since the election of 1832. In strictly arithmetical terms this may have been the case, but could such a comparison have any worth, given that almost every aspect of political life bar the name of the party had been profoundly transformed in the intervening 165 years?

6 States, Societies and Wishful (Emic) Sociological Thinking

Thus, while at any one point in time almost all state frontiers are *etic* (in the sense that it is not difficult to establish ‘objective’ agreement about where they lie, or about the status of competing claims about where they ought to lie, so that categorisation of survey respondents to different state territories is straightforward) social scientists have an unwelcome habit of imposing a series of rather *emic* interpretations of their significance. The most important of these, and one which lies behind the specific problem identified above of category permanence, is the idea, rarely made explicit, but relentlessly followed, that a state is in some sense a self contained ‘society’, such that comparing respondents in different states allows us to compare different ‘societies’.

7 Society as a Category Without Boundaries

By logical definition, society is not a category and has no ‘boundaries’, spatial or temporal, as ‘society’ is, at heart, the process of crossing them. Society is inexorably incontinent and promiscuous. Since the invention of a robust oral tradition or the invention of writing the living have been able to communicate with those long dead, or those so geographically remote or simply numerous that they might never meet them face-to-face. In contrast to its mortal human members each anchored in a mortal physical body that can only ever be in one place at a time (even university professors), ‘society’ is infinite. It is essentially, as Benedict Anderson has put it (1991) an ‘imagined’ community.

That is not to say that all social relations are equal, or that everyone is connected to everyone who has ever lived. Social relations, face-to-face, or face to paper, screen or loud-speaker, cluster and concentrate in time and space. However, the problem comes when we imagine that this cluster and concentration is so great that it neatly corresponds to state boundaries, each relatively hermetic, so that we can therefore make straightforward comparisons across ‘societies’ plural, as if these were sensible categories.

States are increasingly ‘porous’, as illustrated by their ever more energetic but largely unsuccessful attempts to control cross-frontier migration. (Anyone doubting this should compare US expenditure on policing the Mexican border with figures for the numbers crossing it). Visual, aural, oral and bodily communication (books, letters, television, films, radio, newspapers, internet, telephony, private and public transport) are ever faster, cheaper and accessible and not only pay ever less attention to state boundaries, but even escape what control states bother to muster altogether (the internet). State intellectual autarky is theoretically possible (Myanmar, North Korea) but comes at the enormous cost of virtually renouncing any economic or social development. In Europe state interference with such communication

risks courting derision (e.g. ‘Spycatcher’ in the UK) and almost certainly guarantees maximum coverage for the ideas they might attempt to suppress. This is not to deny that, for example, globalised communications corporations such as the Murdoch empire do not shape or even control the news agenda. However any such control hardly coincides with state boundaries. Politicians court Murdoch, not *vice versa*. Economic markets, never totally ‘national’ are now more than ever global. Ideas have always crossed state frontiers, as in the case of the major religions such as Buddhism, Catholicism, Judaism, or Islam.

This requires us to distinguish common sense notions of ‘society’ from scientific ones. The former virtually aligns societies, *plural*, with states. A scientific one would recognise that ‘society’, *singular*, has, always been spatially and temporally incontinent. In a sense we all inhabit distinct ‘societies’ and levels of sociation with concentric boundaries. State frontiers *may* be a significant or relevant boundary within this system, but only one among many.

Finally, once we make the mistake of assuming that distinct states are separate ‘societies’ there is another conceptual cliff off which some ‘comparative’ researchers willingly leap. Lacking either a time machine or reliable longitudinal data, it is tempting to assume that comparisons made across such individual societies at the same or similar points in time can reveal societies, plural, at different stages of evolution along a common historical path of ‘Society’ singular, such that we can read off longitudinal conclusions from transversal comparisons. This basic error is rampant in contemporary social science and I myself have made it many times (e.g. MacInnes 1998b).

The seventeenth century ‘patriarchalist’ Sir Robert Filmer was a supporter of the ‘Divine Right of Kings’ and political opponent of Hobbes. In the course of attacking what he correctly saw to be the mischievously radical and democratic implications of Hobbes’ ideas about *Leviathan* and self interest, he used an astute perception of the difference between a stock and a flow to lampoon the idea that ‘rights’ could ever be ‘natural’. Socially constructed, as opposed to divinely inspired, ‘rights’, identities or social statuses were inherently transversal in character, in contrast to the inherently longitudinal character of generational reproduction:

But where there is an equality by nature, there can be no superior power. There every infant at the hour it is born in, hath a like interest with the greatest and wisest man in the world. Mankind is like the sea, ever ebbing or flowing, every minute one is born another dies. Those that are the people this minute, are not the people the next minute. In every instant and point of time there is a variation. No one time can be indifferent for all mankind to assemble. (Filmer, 1991 [1680]: 142)

Modern social science threw out a rather valuable longitudinal baby with the ancienne regime bathwater when it adopted an implicit model of ‘society’ as an essentially transversal structure. Such a vision might have been required by democratic theory in search of

an ‘indifferent’ time for mankind to exercise its newly imagined rights, and it may have been profoundly strengthened by the nationalist imagination of societies (plural) moving up and down calendrical empty homogenous time (Anderson, 1991), but it is a simplification of the nature of society with various unfortunate results.

8 State Boundaries as a Proxy for Other Categories

Equating state boundaries with societies is not the only problematic interpretation of their meaning and significance. They are also routinely assumed to correspond, more or less, with such categories as: ‘language’; ‘nation’; ‘norms’; ‘values’; ‘laws’; ‘political system’; ‘shared history’; ‘territory’; ‘population’; ‘people’; ‘country’; ‘race’; ‘ethnie’ and ‘society’. There is a further, especially prestigious member of this list: ‘culture’. Comparative social scientists ought to shun this concept wherever possible. It risks becoming an analytic dustbin into which we empty all variance that we cannot understand, or wish to save the effort of analysing, whether across space or time. It is a category that saves us the bother of thinking, but at the risk of condemning our analyses, no matter how advanced in other ways, to unwittingly reproduce banalities. These are deliberately strong words. As an instructive preliminary mental exercise, consider what, if anything, is *not* part of culture? Can we seriously pretend to undertake cross-‘everything’ analyses?

Some of the items in the above list, *some* of the time, in some places, correspond roughly with state boundaries. Much of the time none of them do. Nor is this simply a question of messiness at the margin: a handful of German speakers in France, stateless ‘nations’ within Spain or the UK, administrative decentralisation in federal or quasi-federal states, quaint historical detritus of old empires such as Gibraltar or the Spanish cities in Morocco; fiscal ‘shelters’ like Andorra or Monaco, or micro states like Luxembourg, and so on. Rather it returns us to the more fundamental question of what we think ‘society’ is, and thus the nature of our ‘units’ of analysis. Giving these units different pseudonyms does not solve the problem of what we in fact imagine them to comprise.

States manage their constituent populations. Regardless of their internal social, cultural, political or economic heterogeneity or complexity, from the perspective of the state and its need to secure legitimacy, collect taxes or provide services (health, education, security) this population constitutes a conceptually coherent unit of analysis or ‘category’. There are some contexts in which such a category is also a relevant one for social researchers, e.g. policy analysis, or any analyses where some variable of interest corresponds with the boundary of the state. But just because states (or supra state bodies) fund most research, this should not lead us to assume any identity between their practical objectives and our analytical ones.

9 Reciprocal Failures of Imagination

At the heart of these issues lie two complementary errors in the way we sociologically imagine the nature of our respondents. The first, rather well known one, is the over-socialised conception of the individual made famous by Dennis Wrong over forty years ago (Wrong 1960). In such a vision, the respondent is little more than a cypher for the 'society' (aka population bounded by a state frontier) that the social survey seeks to describe. There is more to individuals than variance that the perfect survey would capture perfectly.

However, I think a rather greater problem facing comparative research, especially in the wake of that celebration of limitless subjectivity known as the 'cultural turn', is what we might call the over-individualised conception of society. The best succinct critique of this has come from a social psychologist. Although in the following excerpt Billig (1995) is discussing the concept of 'identity', his analysis applies perfectly to the argument I am putting forward here about 'society':

The problem starts when one expects to find 'identity' within the body or mind of the individual. This is to look in the wrong place for the operation of identity. ... To have a national identity is to have a way of talking about nationhood. ... only if people believe that they have national identities, will such homelands, and the world of national homelands, be reproduced. ... Nor is national identity to be explored by taking a scale from the psychological library of tests and administering it to a suitable populations. ... National identities are forms of social life, rather than internal psychological states; as such, they are ideological creations, caught up in the historical processes of nationhood.

10 Conflating State Boundaries and Society: (1) Britishness

On 11 March we interrupted our seminar in Madrid to stand for some minutes silence in homage to those massacred on their way to work by the terrorist bombs one year before. No less than one third of those murdered by these bombs were not Spanish nationals. One of the first acts of the government after the tragedy was to confer on them posthumously what they might well have struggled in vain to achieve while alive: Spanish citizenship. The government probably had other than purely altruistic motives for doing so: it made it much easier in the days that followed to speak as if everyone involved in the massacre were in some sense Spanish, the target had been Spain and the Spanish people and that these were also the appropriate object of 'international' solidarity and sympathy. It facilitated the government's role in joining in the public mourning for, and representation of, those victims it might previously have pursued as illegal immigrants. Those massacred on the Madrid commuter trains were certainly all 'Spanish' in the sense that on that fateful day they were in the Spanish state: just as we are today. However to transform them into

symbols of a category 'Spain', as both the bombers and the government did, is to oversimplify a more complex reality. Yet this is precisely the oversimplification we make if we treat respondents in a 'comparative' survey as simple 'representatives' of the 'societies' in which they find themselves. While writing up this paper for publication, bombs exploded in London, murdering a similarly diverse group of Londoners. The alleged suicide bombers, were, conversely, all 'British'. But apart from their place of birth, or rights to a passport, what is the specifically 'British' society that we might think of these bombers as pertaining to?

Senior and experienced politicians in Britain, who might be expected to know better, seem condemned to respond to an irresistible urge to pontificate about Britishness, Englishness and national identity. However this predilection is rather relevant for our purposes as it not only poses the question 'What is Britain?' (or generically 'what *is* a society or the content of a state boundary?') but is also open to some empirical scrutiny as we have extensive survey evidence about what people who happen to be located within the UK state boundary think of Britain and whether they see themselves as 'British'.

Talking about Britishness and (in earlier times) Englishness is an old habit that long predates John Major's risible efforts to crib George Orwell. Speaking to the Annual Dinner of the Royal Society of St George in 1924, Stanley Baldwin spoke of the 'Imperishable scent of old England' and three-quarters of a century after the majority of England's population lived in towns, waxed lyrical about 'the sight of a plough team coming over the brow of a hill, the sight that has been seen in England since England was a land, and may be seen in England long after the Empire has perished and every works in England has ceased to function, for centuries the one eternal sight of England' (Baldwin, 1926). Some civil servant might have told the poor Prime Minister about tractors. More recently the late Robin Cook championed Chicken Tikka as the imperishable scent of multicultural Britain, and Gordon Brown, in his prudent way, declared that the essence of Britishness was that it has no essence, being 'multi-cultural, multi-ethnic and multi-national ... the United Kingdom has always been a country of different nations and thus of plural identities – a Welshman can be Welsh and British just as a Cornishman or woman is Cornish, English and British – and may be Muslim, Pakistani or Afro Caribbean, Cornish, English and British...'. This did not stop him proceeding to give a single definition to precisely that Britishness that he had just asserted could only be plural and diverse! (Brown, 1999; 2004).

I am sure you will have guessed by now where I am going. How is it that normally dexterous (at least with words) politicians are reduced to this sort of nonsense when talking about something that they ought to know about, given that they are supposed to run it? How might comparative analysts be expected to fare better? The answer of course, is to

drop that assumption that the etic ‘UK state frontier’ contains an emic ‘Britain’ or ‘British society’ populated by a universe of ‘British’ prospective survey respondents each carrying around a ‘British’ set of substantive values or way of understanding categories in their heads. National descriptors are an excellent example of our second type of emic category: the logically impossible one defying coherent definition. They exist to legitimate the existing territorial division of state sovereignties in the world (by implying that a society circumscribed by each border is a distinct ‘assembly of mankind’ in Filmer’s terms which, sharing certain ‘values’, inhabits a distinctive state).

Indeed, we have good reason to think that in many contexts many may not think of themselves as British at all. Our evidence on this is also a good example of the way respondent understanding of categories is about much more than issues of translation across different languages. The UK Labour Force Survey (LFS) asks a number of questions about national and ethnic identity. The fact that it is a large scale quarterly survey allows us to merge the results of several successive waves (we would not imagine that respondent’s ethnic or national identities are liable to change over the course of a few months) to produce a sample size large enough to allow us to investigate the differences between relatively small ‘minority ethnic’ groups. In what follows I use the results from 16 successive quarters from 2000 to 2004. I have excluded Northern Ireland (in some senses the most interesting part of the UK for such an exercise) since, following the census, the LFS uses different ‘ethnic’ categories there.

The LFS asks people what they think of as their ‘national identity’. We might imagine that most would answer ‘British’. If we break this down by ethnic group we find that those most likely to give this answer are those who were themselves born, or whose parents were born, in *Pakistan*. ‘White’ British, being rather more likely to choose the term English, are actually the least likely to say they are British. We can alter these results dramatically by simply posing the question a different way. If we present ‘British’ as an ethnic rather than national category, and ask which *ethnic* group people belong to, we find that almost all ‘whites’ except those born overseas, now *do* report thinking of themselves as British. We can get still different results by asking about birthplace and about ‘nationality’. Table 2 summarises these results. Further analysis, not shown here, suggests that different types of respondent interpret questions about their Britishness quite differently. Some take such questions as being about their sense of Britishness *versus* their sense of Englishness, Scottishness or other such concentric spatial identities. Meanwhile, others, especially those not born in the UK, or belonging to minority ‘ethnic’ groups, interpret it as a question about their *nationality* in the sense of rights to possession of a UK Passport. That is to say, within the same language (English) and same state boundary (the UK) different kinds of respondent interpret the meaning of the categories in the same question in different ways.

Table 1 **Proportion reporting British National Identity by ethnic group**

	%	N
Pakistani	69.4	447
Bangladeshi	67.0	176
Black Caribbean	66.4	440
Indian	65.0	758
Other Black	61.5	39
White and Asian	56.5	69
Other Mixed	48.9	45
White and Black Caribbean	48.3	89
White and Black African	45.2	31
Chinese	40.9	154
Other Asian	39.7	224
Black African	38.4	359
British	36.9	40396
Other	30.6	333
Other White	10.2	1747

Source: LFS 2000Q1 – 2004Q1. Excl. Northern Ireland. Ns are weighted

Most concern about ‘comparative’ research centres on how respondents in different ‘national’ (in fact state) contexts will interpret questions differently. This is a problem for all research, comparative or not, since it seems clear that respondents of different kinds within a state may also routinely understand questions in systematically different ways. Rather than simply being a question of proficient language translation (though this is absolutely necessary too) it is a question of supplementing quantitative work with careful enough ethnographic or qualitative research to alert us to social variation within as well as between states and the range of interpretations respondents might bring to the categories used in the questions we ask.

Table 2 'Britishness': Birthplace by nationality by national identity by ethnicity (000's)

Birthplace	Nationality	National Identity	Ethnic Group		Total
			Other	British	
Elsewhere	Other	Other	1912	141	2053
		British	183	88	271
	British	Other	363	343	706
		British	981	641	1622
Britain	Other	Other	153	511	664
		British	10	106	116
	British	Other	715	24329	25044
		British	770	14072	14842
Total			5087	40231	45318

Source: LFS 2000Q1 – 2004Q1. Excl. Northern Ireland. Ns are weighted

11 Conflating State Boundaries and Society: (2) Variety Within States

Political parties often do not operate across the whole of a state territory but represent regional or peripheral-national interests specific to a sub state area. Failure to account for this properly can lead to misleading results. The following example is from Hakim (2003: 195-6):

A strong regional identity does not necessarily inform voting behaviour. The question asking about people's vote in the 1996 elections listed a large number of nationalist regional parties, but only 8% of respondents (9% after excluding refusals) voted for them. Minority party voters are found in all age groups.

Surprisingly, there is no association between education and political ideology in Spain, in the younger and older generations, among men and women. The least-educated groups are just as likely to be left-wing as the university graduates. The Political ideology and nationalism have no impact at all on ideal family model. ... political ideology has some impact on work orientations (Table 7.4). Supporters of the two main political parties are undifferentiated, with work centrality high among men and low among women, as usual. However, work centrality rises sharply among women and falls sharply among men who vote for the minority parties...

This analysis would be fine, were it not for two things. First, only a small percentage of Spain's population can vote for these minority parties as they are spatially concentrated.

Where they do operate they often command a majority or near majority of votes, as in Catalonia, the Basque Country, Navarre and the Canaries. Indeed in Catalonia the only 'state wide' party that receives votes is the Partido Popular (PP) but with a very low share – around 10-15%. Second, the characteristics of these parties and their electorate is totally heterogeneous. The UPN is a right wing regionalist party that is effectively the PP in Navarre. Herri Batasuna (now illegal) was the political wing of ETA. To analyse them together is hardly sensible. Taking account of this in the analysis would alter Hakim's mistaken results.

12 Some Brief Conclusions

Theory is all very well but what methodological implications flow from this argument? First, we ought to question the distinctiveness of what is currently thought of as 'comparative' research. All research is comparative, and quantitative research that is not alert to how different kinds of respondent (however defined) might understand the same question in systematically different ways is unlikely to produce robust results. Understanding questions is more than a matter of language comprehension. We should be prepared to address 'non' comparative research in just as critical a spirit, for it logically follows that all such research must fall foul of the problems discussed above insofar as it restricts itself, without sufficiently clear and established cause, to the study of social relations within a particular state frontier. 'National' research that does not look beyond state frontiers in the course of its analysis must become steadily more anachronistic in an ever more globalised world.

Second, we ought to question the assumption that state boundaries accurately circumscribe other aspects of social relations that we might want to compare. In particular, it is highly unlikely that they bound empirically distinct and comparable self-contained 'societies' whose functioning we might wish to compare. This implies that the kind of comparisons we might wish to make should be made as specifically as possible: for example referring to particular institutions and seeking to identify the practical spatial boundary of their remit rather than assuming that this stops at the state frontier. One way of improving the 'comparability' of research across boundaries is to marry qualitative and quantitative research more effectively, such that the former provides us both with 'categories' whose symbolic content either remains constant across boundaries or changes systematically in known and predictable ways, and with better knowledge of the diverse and overlapping social boundaries, in addition to state frontiers, that our fieldwork will cross.

Third we should be extremely cautious about treating respondents from different states as representing homogeneous 'values' or 'cultures' that are assumed to typify these states. Fourth, we should use words like 'culture' with care, and drop them where we cannot specify more accurately what we mean by such a catch-all terminology.

References

- Anderson, B. (1991). *Imagined Communities* (2nd. ed). London: Verso.
- Anderson, P. (1978). *Lineages of the Absolutist State*. London: NLB.
- Baldwin, S. (1926). *On England*. Harmondsworth: Penguin.
- Billig, M. (1995). *Banal Nationalism*. London: Sage.
- Bloch, M. (1989). *Federal Society*. London: Routledge.
- Brown, G. (1999). *Speech to Smith Institute*. 15 April. (Available on HMG Treasury website).
- Brown, G. (2004). *British Council Annual Lecture*. 7 July. (Available on HMG Treasury website).
- Craib, I. (1998). *Experiencing Identity*. London: Sage.
- Goffman, E. (1963). *Stigma*. Harmondsworth: Penguin.
- Hakim, Catherine. (2003). *Models of the Family in Modern Societies*. Aldershot: Ashbury.
- MacInnes, J. (1998a). Analysing patriarchy, capitalism and women's employment in Europe. *Innovation*, 11(2), 227-248.
- MacInnes, J. (1998b). *The End of Masculinity*. Buckingham: Open University Press.
- Pike, K. L. (1954). *Language in relation to a unified theory of the structure of human behavior*. Glendale, CA: Summer Institute of Linguistics.
- Wrong, D. (1960). The over-socialised conception of man in sociology. *American Sociological Review*, 26(2), 183-193.

ABOUT THE CONTRIBUTORS

Michael Blohm is a researcher at the European Centre for Cross-Cultural Survey at ZUMA, Mannheim, Germany. He is one of a small team responsible for the research programme, questionnaire development and implementation of the German General Social Survey (ALLBUS). Currently working on his doctorate, Michael is mainly interested in survey methodology, especially non-response and interviewer behavior.

Patricia L. Goerman is a Postdoctoral Research Fellow in the Statistical Research Division at the United States Census Bureau. Her current research focuses on the development and pre-testing of multilingual survey instruments with a focus on Spanish language materials. She received her Ph.D. in Sociology from the University of Virginia in 2004. Her dissertation is a study of Latino immigrants settling in non-traditional areas in the Southeastern United States. Her past research has included in-depth interviews with large numbers of Hispanic immigrants. She holds an M.A. in Sociology from the University of Virginia and a B.A. in Spanish, English, and Latin American and Iberian Studies from the University of Wisconsin-Madison.

Janet A. Harkness is a cross-cultural survey methodologist with expertise in instrument design, adaptation, and implementation. She has a multidisciplinary academic background and training – comparative medieval studies (M.A.); linguistics, cultural anthropology (Ph.D.); survey methods and cross-

national survey research (ZUMA). Before being appointed to the Donald and Shirley Clifton Chair in Survey Science at the University of Nebraska-Lincoln in autumn 2005, she was Director of the International Social Survey Programme for Germany and Senior Scientist at ZUMA. She is a member of the Central Co-ordinating Team of the European Social Survey (ESS) and heads the Translation Work Package for the ESS. She continues her affiliation with ZUMA as a Senior Research Scientist. She has published, presented, and taught on numerous aspects of cross-cultural survey research and communication. Her current projects focus on oral translation in surveys, questionnaire design for comparative projects and CSDI guidelines for cross-national survey research.

Jürgen H.P. Hoffmeyer-Zlotnik studied Sociology at Berlin and Hamburg. He received his doctorate from the University of Hamburg and a “habilitation” from the universities of Cologne and Giessen. First project director at GEWOS, Hamburg (1976/77) then scientific assistant at the University of Trier (1978-1980), he moved to ZUMA in 1980, where he is currently senior project consultant. Main areas of research: (1) standardization and harmonization of demographic and socioeconomic variables in national and cross-national comparison; (2) human ecology, social spatial differentiation of neighborhoods as background variables.

Lilli Japiec has been a survey methodologist at Statistics Sweden since 1989. For the past eight years she has been with the R&D Department. Areas she has worked in include Labour Force Surveys and salary and wages statistics; she also spent six months as a national expert for Sweden at the Director General's Office at Eurostat. Research areas of particular interest include data collection, interview surveys, non-response, and quality management. Lilli recently completed a doctoral thesis on interview quality issues. She is currently scientific secretary of the International Association of Survey Statisticians.

Brian Kleiner is a senior researcher at Westat, a social science research firm outside of Washington, D.C. Trained as a discourse analyst, his current work in part focuses on best practices in survey translation, with consideration of culturally appropriate norms of communication. His other work involves the design and implementation of national surveys and evaluation studies on education-related topics and programs.

Achim Koch is a Senior Scientist at ZUMA, Mannheim, Germany. He came to ZUMA as a social scientist in 1986 and was appointed Director of the German General Social Survey (ALLBUS) in 1995, a position he held until 2004, when he moved to become Senior Scientist in the European Centre for Cross-Cultural Surveys at ZUMA. He is a member of the European Social Survey (ESS) Central Co-ordination Team, and heads the contract monitoring work package for the ESS. His major methodological research interests are non-

response in international contexts and cross-cultural aspects of data collection. Recent publications include several articles on attitudes towards foreigners, articles on field work and non-response, and co-editorship of two volumes, one on political participation and another on social change in Germany.

Lars Lyberg is a Chief Scientist at Statistics Sweden. Formerly he was Head of the Statistical Research Unit (1983-1987), Head of the International Research Unit (1988-1992), and Head of the R & D Department (1993-2000) at Statistics Sweden. He received his Ph.D. from Stockholm University in 1981. Since 1985 he has been the Chief Editor of the Journal of Official Statistics. He is chief editor of Survey Measurement and Process Quality (Wiley, 1997) and co-editor of Telephone Survey Methodology (Wiley, 1988) and Measurement Errors in Surveys (Wiley, 1991) and co-author of Introduction to Survey Quality. His main research interests are non-sampling errors in surveys, data quality, and organizational quality.

Peter Lynn is Professor of Survey Methodology at the University of Essex and was previously Director of the Survey Methods Centre at the National Centre for Social Research. Peter specialises in all aspects of quantitative data collection methodology and has published widely on topics including survey non-response, weighting, data collection mode effects, respondent incentives, advance letters, sample design and survey quality. His full CV is available at www.iser.essex.ac.uk. Currently, Peter is associate director of the UK Longitudinal Studies Centre. Current research

interests encompass measurement error, non-response error and sample design – especially in longitudinal contexts. Peter is editor-in-chief of the on-line journal ‘Survey Research Methods’.

John MacInnes is Reader in Sociology at the University of Edinburgh and Researcher at the Centre d’Estudis Demogràfics, Universitat Autònoma de Barcelona. His chief substantive interests are in gender; ‘reproduction’ and the labour market – family relation; and national identity. Recent books include *Stateless Nations in the 21st Century* [edited with David McCrone, Edinburgh: USGS, 2001] and *The End of Masculinity* [Open University Press, 1998]. Recent papers cover topics such as Work Life Balance in Europe; Dynamic ECHP survey evidence on changes in satisfaction with leisure time following childbirth; Comparisons of national identity in Scotland and Catalonia; The conceptualisation of demographic change in the twentieth century; and the role of the mass media in the UK in reproducing national consciousness. He has used a variety of ‘quantitative’ and ‘qualitative’ methodologies, from event history analysis to psychoanalysis. His particular interest in comparative research turns on the clarifying what, if anything, differs between comparison across respondents from different states (so-called ‘comparative’ research) and routine ‘comparison’ across cases, values and variables in sociological enquiry.

Yuling Pan Yuling Pan is a sociolinguist who works in the Statistical Research Division of the U.S. Census Bureau. She has conducted research in survey translation and pretesting of translation in multiple languages. Her research interests also include cross-cultural communication, language and cultural issues in survey design and implementation. She has co-authored the U.S. Census Bureau Translation Guidelines and is the author and co-author of two books and numerous journal articles in the field of sociolinguistics.

Uwe Warner studied sociology and education at the University of Heidelberg. He has been a senior researcher at CEPS/INSTEAD, Luxembourg since 1989. His main research interests are cross-national comparative survey research and poverty, income and well-being. He was responsible for collecting the Luxembourg data for the European Community Household panel (ECHP). He is the National Co-ordinator of ESS in Luxembourg and the Luxembourg delegate to CESSDA.

ZUMA-Nachrichten Spezial

Die Reihe ZUMA-Nachrichten-Spezial dient dazu, den Forschungsstand größerer Arbeits- oder Forschungsbereiche bei ZUMA zu dokumentieren oder die Ergebnisse von Konferenzen und Symposien vorzustellen (http://www.gesis.org/publikationen/Publikationen/zeitschriften/ZUMA_Nachrichten_spezial/). Bisher sind elf Bände erschienen.

* * *

ZUMA- Nachrichten Spezial Band 1 (vergriffen)

Text Analysis and Computers

**Hrsg. von Cornelia Züll, Janet Harkness und Jürgen H.P. Hoffmeyer- Zlotnik
Mannheim, ZUMA, 1996, 132 Seiten, ISBN 3-924220-11-5**

Das Heft entstand im Anschluß an eine internationale Tagung zur computerunterstützten Textanalyse, bei der sich Wissenschaftler aus den verschiedensten Disziplinen trafen. Die hier abgedruckten Papiere der eingeladenen Hauptredner dokumentieren den Forschungsstand in vier Bereichen: Computer-Assisted Content Analysis: An Overview (*E. Mergenthaler*); Computer-Aided Qualitative Data Analysis: An Overview (*U. Kelle*); Machine-Readable Text Corpora and the Linguistic Description of Language (*Chr. Mair*); Principle of Content Analysis for Information *Retrieval* (*J. Krause*). Der Band ist auch als PDF-Datei im Internet verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 2 (vergriffen)

Eurobarometer. Measurement Instruments for Opinions in Europe

**Hrsg. von Willem E. Saris und Max Kaase
Mannheim: ZUMA 1997, ISBN 3-924220-12-3**

In der Empirischen Sozialforschung finden in Europa Telefoninterviews anstelle von face to face-Interviews zunehmende Verbreitung. Im Rahmen der zweimal jährlich für die Europäische Kommission in Brüssel durchgeführten Repräsentativbefragungen in den Mitgliedsländern der Europäischen Union, den sogenannten Eurobarometern, ergab sich für die Erhebung vom Frühjahr 1994 (EB 41.0) die Möglichkeit, durch eine zeitgleich mit einem weitgehend identischen Fragenprogramm stattfindende Telefonbefragung in den damaligen zwölf Mitgliedsländern der EU, systematisch Effekte der unterschiedlichen Stichprobenansätze und Erhebungsmethoden zu untersuchen. Dabei konnte das Analysespektrum noch durch eine Telefon-Panelkomponente in dreien der zwölf EU-Länder für das face to face-Eurobarometer erweitert werden. Die Beiträge im vorliegenden Buch untersuchen auf dieser Grundlage methodische und methodologische Fragestellungen, die insbesondere für die international vergleichende Sozialforschung, aber auch für die Markt- und Meinungsforschung in Europa von großer Bedeutung sind. Der Band ist auch als PDF-Datei im Internet verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

ZUMA-Nachrichten Spezial Band 3
Cross-Cultural Survey Equivalence.

Hrsg. von J. Harkness

Mannheim: ZUMA 1998, 187 Seiten, ISBN 3-924220-13-1

This volume, the third in the ZUMA-Nachrichten-Spezial series on methodological issues in empirical social science research, is devoted to issues of cross-cultural methodology. The focus is on issues of equivalence, the key requirement in cross-national and cross-cultural comparative research. As the contributions indicate, equivalence is, however, better thought of in terms of equivalencies - in social science surveys and in other standardised instruments of measurement. Contributors come from different countries and continents and from widely differing research backgrounds, ranging from linguistics to survey research and its methodologies, to cultural anthropology and cross-cultural psychology. They are: Timothy P. Johnson, Fons J.R. van de Vijver, Willem E. Saris, Janet A. Harkness and Alicia Schoua-Glusberg, Michael Braun and Jacqueline Scott, Ingwer Borg, Peter Ph. Mohler, Tom W. Smith and Janet A. Harkness. This volume can be downloaded as a PDF file (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/)

* * *

ZUMA-Nachrichten Spezial Band 4 (vergriffen)

Nonresponse in Survey Research

Hrsg. von A. Koch und R. Porst

Mannheim: ZUMA 1998, 354 Seiten, ISBN 3-924220-15-8

This volume, the fourth in the ZUMA-Nachrichten Spezial series on methodological issues in empirical social science research, takes up issues of nonresponse. Nonresponse, that is, the failure to obtain measurements from all targeted members of a survey sample, is a problem which confronts many survey organizations in different parts of the world. The papers in this volume discuss nonresponse from different perspectives: they describe efforts undertaken for individual surveys and procedures employed in different countries to deal with nonresponse, analyses of the role of interviewers, the use of advance letters, incentives, etc. to reduce nonresponse rates, analyses of the correlates and consequences of nonresponse, and descriptions of post-survey statistical adjustments to compensate for nonresponse. All the contributions are based on presentations made at the '8th International Workshop on Household Survey Nonresponse'. The workshop took place in September 1997 in Mannheim, Germany, the home base of the workshop host institute, ZUMA. Twenty-nine papers were presented and discussed, of which twenty-five are included here.

ZUMA-Nachrichten Spezial Band 5
A review of software for text analysis
Alexa Melina & Cornelia Zuell

Mannheim: ZUMA 1999, 176 Seiten, ISBN 3-924220-16-6

The book reviews a selection of software for computer-assisted text analysis. The primary aim is to provide a detailed account of the spectrum of available text analysis software and catalogue the kinds of support the selected software offers to the user. A related, more general, goal is to record the tendencies both in functionality and technology and identify the areas where more development is needed. For this reason the presented selection of software comprises not only fully developed commercial and research programs, but also prototypes and beta versions. An additional aspect with regards to the kinds of software reviewed is that both qualitative and quantitative-oriented types of research are included. Depending on research purposes and project design the text analyst can profit from available tools independently of their orientation. The following fifteen programs are reviewed: AQUAD, ATLAS.ti, CoAN, Code-A-Text, DICTION, DIMAP-MCCA, HyperRESEARCH, KEDS, NUD*IST, QED, TATOE, TEXTPACK, TextSmart, WinMAXpro, and WordStat and the criteria and methodology used for selecting them are delineated. Der Band ist auch als PDF-Datei im Internet verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 6
Sozialstrukturanalysen mit dem Mikrozensus
Hrsg. von Paul Lüttinger

Mannheim: ZUMA 1999, 402 Seiten, ISBN 3-924220-17-4

Im Oktober 1998 veranstaltete die Abteilung Mikrodaten von ZUMA die Konferenz "Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt", an der vorwiegend Nutzer des Mikrozensus teilnahmen. Hauptziel dieser ersten Nutzerkonferenz war es, ein Forum für den Informationsaustausch zwischen den Datennutzern und den statistischen Ämtern zu schaffen. Die mehr als 20 Vorträge gingen deutlich über die von den statistischen Ämtern veröffentlichten Standardergebnisse zum Mikrozensus hinaus und sind weitgehend in diesem Band ZUMA-Nachrichten Spezial abgedruckt. Die Autoren sind: Walter Müller; Karl Brenke; Esther Hansch und Michael-Burkhard Piorkowski; Friedhelm Pfeiffer; Jürgen Schupp, Joachim Frick, Lutz Kaiser und Gert Wagner; Elke Wolf; Dietmar Dathe; Bernd Eggen; Erich Stutzer; Carsten Baumann; Susanne von Below; Thomas Bulmahn; Martin Groß; Reiner H. Dinkel, Marc Luy und Uwe Lebok sowie Wolfgang Strengmann-Kuhn. Der Band ist als PDF-Datei im Internet verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

ZUMA-Nachrichten Spezial Band 7
Social and Economic Analyses of Consumer Panel Data
Georgios Papastefanou, Peter Schmidt, Axel Börsch-Supan,
Hartmut Lüdtke, Ulrich Oltersdorf (Eds.)
Mannheim: ZUMA 2001; 212 Seiten; CD-Rom

Eine von der Abteilung Einkommen und Verbrauch von ZUMA organisierte Arbeitsgruppe hat sich mit datentechnischem Handling und Analysepotential von komplexen Verbraucherpaneldaten, am Beispiel des ConsumerScan Haushaltspanels der Gesellschaft für Marktforschung (GfK, Nürnberg) beschäftigt und die Ergebnisse in einem Symposium im Oktober 1999 vorgestellt. Die überwiegende Zahl der vorgetragenen Arbeiten, die man als Werkstattberichte ansehen kann, sind in diesem Band abgedruckt. Neben einem detaillierten Einblick in die Praxis und das Datenerhebungsprogramm von Verbraucherpanels, wie sie z.B. bei der Marktforschungen der GfK unterhalten werden, enthält der Band z.B. Untersuchungen zu Fragen der Flexibilität von Preisbildungsvorgängen, des Lebensstils im alltäglichen Konsums, der Gesundheitsorientierung im Konsumverhalten, der Umweltorientierung und ihrer Umsetzung im Kauf alltäglicher Haushaltsprodukte. Der Band enthält eine CD-ROM mit Dokumenten und Codebüchern der aufbereiteten ZUMA-Verbraucherpaneldaten 1995. Der Band ist auch als PDF-Datei im Internet verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 8
Von Generation zu Generation
Hrsg. von Jan van Deth
Mannheim: ZUMA 2002, 68 Seiten, ISBN 3-924220-23-9

Aus Anlass der Ehrung von Prof. Dr. Max Kaase, Prof. Dr. Walter Müller und Prof. Dr. Hansgert Peisert für ihre langjährige und richtungsweisende Mitarbeit in der Mitgliederversammlung des ZUMA e.V. fand am 14. Juni 2002 eine wissenschaftliche Tagung statt. Der Band enthält Beiträge von Jan van Deth, Hubert Feger, Jürgen Rost, Erwin K. Scheuch, Andreas Diekman und Hans-Dieter Klingemann. Die Beiträge sind auch online verfügbar (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 9

QUEST 2003

Questionnaire Evaluation Standards

Peter Prüfer, Margrit Rexroth, Floyd Jackson Fowler, Jr. (Eds.)

Mannheim: ZUMA 2004, 216 Seiten, ISBN 3-924220-27-1

This volume, the ninth in the ZUMA-Nachrichten Spezial series on methodological issues in empirical social science research takes up issues of question and questionnaire evaluation. The papers in this volume discuss practical as well as theoretical aspects of questionnaire evaluation. All contributions are based on presentations made at the fourth QUEST (Questionnaire Evaluation Standards) conference which took place from October 21 - 23, 2003 at ZUMA in Mannheim. There were 26 attendees from 9 countries representing 14 organizations: Bureau of Labor Statistics, USA, Center for Survey Research, University of Massachusetts, USA, Institut für Demoskopie Allensbach, Germany, National Center for Health Statistics, USA, National Center for Social Research, U.K., Office of National Statistics, U.K., Statistics Canada, Statistics Finland, Statistics Netherlands, Statistics New Zealand, Statistics Norway, Statistics Sweden, U.S. Census Bureau, ZUMA, Germany. This volume can be downloaded as a PDF file (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 10

Theory, Laws and Measurement

Michael Braun & Peter Ph. Mohler (Eds.)

Mannheim: ZUMA (forthcoming) ISBN 3-924220-28-X

* * *

ZUMA-Nachrichten Spezial Band 11

Methodological Aspects in Cross-National Research

Jürgen H.P. Hoffmeyer-Zlotnik & Janet A. Harkness (Eds.)

Mannheim: ZUMA 2005, 305 Seiten, ISBN 3-924220-29-8

The idea for this volume was born during the Sixth International Conference on Social Science Methodology in Amsterdam in August 2004, organised by the International Sociological Association Research Committee 33 on Logic and Methodology. Most of the contributions in this volume are proceeding papers from the Amsterdam conference.

The contributions in this volume are organised in four parts. The first part deals with designing and implementing cross-cultural surveys. The second part consists of three papers that deal with different issues of comparability or “equivalence”. The third part of

the volume brings together papers on with harmonising socio-demographic information in different types of surveys. The last section of the volume contains papers that discuss individual socio-demographic variables in cross-national perspective. This volume can be downloaded as a PDF file from December 2007 on (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).

* * *

ZUMA-Nachrichten Spezial Band 12
Conducting Cross-National and Cross-Cultural Surveys
Papers from the 2005 Meeting of the International Workshop on
Comparative Survey Design and Implementation (CSDI)
Janet A. Harkness (ed.)
Mannheim: ZUMA 2006, 123 Seiten, ISBN 3-924220-31-X

The papers in this volume stem from the third annual meeting of the International Workshop on **Comparative Survey Design and Implementation (CSDI)**. Initiated in 2002, the Workshop developed out of cross-cultural symposia held at ZUMA throughout the nineteen nineties. One of CSDI's primary goals is to promote research into methodological issues of particular and sometimes unique salience for cross-cultural and cross-national survey research. For more information visit the CSDI website (www.csdi-workshop.org).

The seven papers are good illustrations of the broad spectrum of research fields in which CSDI researchers are engaged. The volume begins and ends with two framework papers, the first discussing what makes cross-national research special, the last on where we begin to draw boundaries between entities to be compared in "comparative" research. The five remaining papers discuss (in order of the volume): the rich information available from the multinational European Social Survey on data collection; socio-demographic measurement and comparability in the cross-national context, again with reference to the European Social Survey; cognitive pre-testing of translated questionnaires; communicative issues across cultures in telephone interviews; and preliminary work on guidelines on using interpreters underway at the U.S. Census Bureau. The last-mentioned papers reflect research concerns in U.S. cross-cultural contexts. This volume can be downloaded as a PDF file from March 2007 on (http://www.gesis.org/publikationen/zuma_nachrichten_spezial/).